DOCUMENT RESUME

ED 218 921 FL 012 805

AUTHOR Swinton, Spencer S.; Powers, Donald E.

TITLE Factor Analysis of the Test of English as a Foreign Language for Several Language Groups. TOEFL Research Reports, 6.

INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-80-32
PUB DATE Dec 80
NOTE 88p.

EDRS PRICE MF01/PC04 Plus Postage.

DESCRIPTORS Arabs; Educational Research; *English (Second Language); Factor Analysis; *Factor Structure; Foreign Students; Higher Education; *Language Tests; Spanish Speaking; *Test Interpretation; Young Adults

IDENTIFIERS Africans; Chinese People; Germans; Iranians; Japanese People; Scandinavians; *Test of English as a Foreign Language

ABSTRACT

A factor analytic study of the Test of English as a Foreign Language (TOEFL) was undertaken to determine the component abilities that underlie performance on the test for several major language groups: African, Arabic, Chinese (non-Taiwanese), Farsi, Germanic, Japanese, and Spanish. Evidence was found that three major factors underlie performance on the TOEFL and that these factors are relatively unambiguous in their interpretation. A factor underlying the listening comprehension section was noted for each language group; however, there were differences among the language groups in the interpretation of two of the factors. The African, Arabic, Chinese, and Japanese groups were generally similar on a factor underlying performance on structure, written expression, and reading comprehension items; and on another separate factor underlying vocabulary items. The Spanish and Germanic groups were also similar on each of two other factors, which correspond to the TOEFL subscores (structure/written expression, and reading comprehension/vocabulary). The vocabulary factor exhibited positive correlations with age and degree-intentions in nearly every language group, suggesting that vocabulary is the most likely of any of the abilities to develop with training or experience. Implications for the interpretation of TOEFL subscores are discussed. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218921

TOEFL Research Reports

REPORT 6
DECEMBER 1980

FACTOR ANALYSIS OF THE TEST OF ENGLISH AS A FOREIGN LANGUAGE FOR SEVERAL LANGUAGE GROUPS

Spencer S. Swinton
Donald E. Powers

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational
Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TOEFL
Educational Testing Service

FL012805

The Test of English as a Foreign Language (TOEFL) was developed in 1963 by a National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of over thirty organizations, public and private, that were concerned with testing the English proficiency of non-native speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program and in 1973 a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; Graduate Record Examinations Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the Graduate Record Examinations Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

Factor Analysis of the Test of English as a Foreign Language for Several Language Groups

Spencer S. Swinton

Donald E. Powers

Educational Testing Service

Princeton, N.J.

RR 80-32

Copyright © 1980 by Educational Testing Service. All rights reserved.
Unauthorized reproduction in whole or in part is prohibited.

Abstract

A factor analytic study of the Test of English as a Foreign Language (TOEFL) was undertaken to determine the component abilities that underly performance on the test for several major language groups. Alternative analytic procedures were employed that included (a) computing separate factor solutions for each language group, (b) rotating separate solutions to a common target, and (c) examining the relationships between factors and a variety of candidate background variables.

The structure of the test is discussed in terms of the similarities and differences found in each language group. The implications of the results for test development, for the interpretation of TOEFL subscores, and for English language training are presented.

Introduction

The present study was undertaken to provide additional evidence of the construct validity of the Test of English as a Foreign Language (TOEFL) by determining precisely what component abilities the test measures, i.e., the explanatory constructs that account for examinee performance. One general technique for doing so is factor analysis, a set of analytic methods that provides an estimate of the amount and nature of test variance that is attributable to response consistencies as these are reflected in interitem correlations.

This factor analytic study of the TOEFL was intended to add to our knowledge of the test. An accumulation of knowledge is considered important, since evidence of a measure's construct validity is generally thought to result not from a single study, but from the triangulation made possible by a variety of related research (American Psychological Association, 1974). Previous studies that have contributed to the understanding of what the test measures are those by Angelis, Swinton, and Cowell (1979); Angoff and Sharon (1970); Clark (1977); and Pike (1979).

The present study was also intended to provide information about minor dimensions of test performance that, although not to be interpreted as psychological constructs, might be useful in defining molecular dimensions of test-score variance. Such dimensions might be highly relevant to sharpening test-content specifications and to guiding the construction of additional parallel forms of the test, in much the same manner as a previous factor analysis of the GRE Aptitude Test helped to define additional relevant item-classification dimensions for that test (Powers & Swinton, 1976).

Finally, since TOEFL is by nature an international test administered to an extremely heterogeneous group of examinees, it seemed desirable to determine the extent to which the test measures the same constructs for various language groups, an assessment that was thought to have important implications for the interpretation of test scores.

Methods

Sample Selection

Samples of examinees ranging in size from approximately 600 to 1,000 from each of seven major language groups were selected from the total population of TOEFL candidates who registered for the November 1976 international administration. The language groups that were studied included African, Arabic, Chinese (non-Taiwanese), Farsi, Germanic, Japanese, and Spanish.

The choice of language groups for analysis was dictated by both theoretical and practical considerations. As in the earlier study by Angoff and Sharon (1970), the languages were to be (1) as diverse as possible, representing different branches of the Indo-European family and several different non-Indo-European language groups, and (2) represented by a sufficiently large number of candidates taking a particular TOEFL form, in this case form YTF4.

The large volume of Chinese- and Farsi-speaking TOEFL candidates suggested a need to include these two groups. Japanese and Arabic, besides representing large volume, provided theoretically interesting contrasts with the first two languages. Among African language groups, the Guinea-Senegalese group, including Yoruba, Ibo, Efik, and other West African languages, offered a sufficiently large sample for analysis. Although the agglutinative Bantu languages were of considerable theoretical interest, they were not represented by a sufficient number of candidates to be included. Among European languages, Spanish represented the largest candidate volume, and Germanic languages (German, Dutch, Swedish, Norwegian, Danish, and Icelandic), although relatively low in volume, were theoretically important because of their close linguistic affinity to English. It was recognized that considerable variety exists within the Guinea-Senegalese and Germanic language families, but the relatively small number of candidates in each language made pooling necessary. Thus, the resulting set of languages contained five of the six examined by Angoff and Sharon (1970). We substituted Farsi for Gujarati among Indo-European languages because of changes in candidate volume and added the Guinea-Senegalese language group to the non-Indo-European set. The language groups and samples were identical to those used by Alderman (1980) in a study of item performance across language groups. Thus comparisons with that study are possible.

Analyses

Each of the 149* items in the TOEFL was used as a variable in several factor analyses. For each language group, the matrix of tetrachoric interitem correlations was analyzed. Items passed by 96% or more of a group were omitted in order to avoid convergence problems with the tetrachoric computation algorithm. Initial correlation estimates were based on a method devised by Tucker and were used as starting values for minres solutions (Harman, 1976). Communality estimates were obtained as a byproduct of the iterative minres procedure. By examining the break in the magnitude of eigenvalues and subsequently rotating several numbers of factors, according to the varimax criterion we were able to choose a

---

* One structure item, number 13 in Section II, was not included in the operational scoring, because of British/American vocabulary differences, and was correspondingly omitted from this analysis.

provisional orthogonal minres solution for each language group. On the basis of the variations in structure suggested by each group's solution, we specified a common target matrix for all groups. We then performed a rotation to this target for each group to insure optimal comparability while placing minimal constraints on the emergence of the kinds of structural differences that had emerged from the within-group varimax rotations.

When simple structures had been estimated by the above method, candidate background variables were regressed on factors through factor extension analysis (Dwyer, 1937) in order to determine the relationship of each factor to various candidate background characteristics (e.g., sex, age, degree planned, intended major field, number of times TOEFL taken, native country, and reason for taking TOEFL). This technique provided further interpretive information, allowing examination of the relationship of these extension variables to the factors underlying test performance, without affecting the factor structure determined for the test.

Results

Description of Language Groups

Table 1 shows the number in each language group on which the reported analyses are based, and the average number of items answered correctly. As is clear, there is substantial variation among the language groups. For example Germanic speakers received higher scores than any other group on each subtest; Farsi speakers received lower scores than any other group. Furthermore, the subtest profiles are different for the groups: group performances are not necessarily uniformly high or low on all subtests. For example, although African speakers rank second on Structure and Written Expression, they rank only sixth on Listening Comprehension. These between-group differences in the level of English proficiency have implications for the interpretation of factors and the comparison of structures across language groups. Variation of item difficulty across language groups within a subtest is *prima facie* evidence of multidimensionality of that subtest. Alderman's (1980) study of item difficulties reveals that such departure from unidimensionality exists for these samples and for this form of the TOEFL.

Preliminary Analyses

Inspection of eigenvalues suggested that different numbers of factors might be appropriate for each of the language groups. With the exception of the Germanic group, three, four, or five factors appeared appropriate for each language group. Trial rotations of these numbers of factors were carried out. For the Germanic group, however, the size of

Table 1

Mean Number of Items Correct and Standard Deviations for Seven Language Groups for Each TOEFL Subtest

Language Group	N		TOEFL Scores: Listening Comprehension	Structure and Written Expression	Reading Comprehension and Vocabulary
African	740*	Mean	29.5	28.8	41.2
		S.D.	8.5	6.1	9.2
Arabic	686	Mean	32.4	23.0	35.0
		S.D.	9.6	6.9	10.5
Chinese	998	Mean	31.5	24.8	37.3
		S.D.	8.9	6.4	9.9
Farsi	987	Mean	27.6	19.4	29.6
		S.D.	9.8	6.5	9.3
Germanic	596*	Mean	43.6	30.7	48.0
		S.D.	5.5	5.2	7.1
Japanese	997	Mean	30.6	22.8	36.3
		S.D.	8.6	6.3	10.1
Spanish	991	Mean	35.5	24.7	44.8
		S.D.	9.8	7.3	9.0

*One composition of the combined language groups was as follows:

Language	n
African	
Efik	68
Ibo	413
Fanti	40
Yoroba	219
Germanic	
Danish	39
German	314
Dutch	157
Swedish	86

eigenvalues suggested that as many as eight factors might be important for a good fit to the item intercorrelations or a comprehensive description of the group's test performance. This greater complexity for the Germanic group suggests that TOEFL is tapping a larger set of differentiated abilities in this group. One reason for this result may be that, because of the overall high level of English proficiency encountered in the Germanic speakers, their responses included less noise attributable to guessing, making it possible for a larger number of minor dimensions of covariation to appear consistently among items (but see the Discussion [p. 13] for a more substantive hypothesis).

Table 2 gives the numbers of items from each subsection with highest loadings on each factor. These are the results of the separate varimax solutions for each language group. Those cases in which one-half or more of the items of a given type load on a single factor are asterisked.

Inspection of Table 2 reveals that the great majority of Listening Comprehension items load on a single factor in each language group. Furthermore, this factor shows relatively few loadings from other item types in most language groups, suggesting that in general, the Listening Comprehension items appear to determine a distinct factor, as would be expected from their unique aural, rather than written, mode of presentation.

The Structure and Written Expression grouping of items tend to load on a common factor for the African, Chinese, and Japanese groups; but for the Arabic, Farsi, and Spanish groups, Written expression determines a relatively independent factor. For the Chinese and Total groups--and to a lesser extent, for the African, Germanic, and Japanese--Written Expression clusters with Reading Comprehension.

The other operational grouping, Vocabulary and Reading Comprehension, forms an empirical cluster for Arabic and Spanish speakers. It also exhibits a moderate tendency to cluster for the African and Farsi groups and a weaker one for the Germanic speakers in these independent varimax rotations. Vocabulary determines a relatively independent factor for the Chinese, Japanese, and Total Groups, however, and exhibits some tendency to split from Reading Comprehension in the African-, Germanic-, and Farsi-speaking samples.

To summarize the varimax rotations, we found that Listening Comprehension appears to be a univocal factor in all groups, although Structure items tend to load somewhat on this factor in the Spanish, Germanic, Farsi, and Arabic groups. Structure and Written Expression items form a clear cluster for the African, Chinese, and Japanese groups, but Written Expression separates from Structure to form a distinct factor in the Arabic and Farsi groups. Vocabulary and Reading Comprehension form a relatively clear cluster in the Arabic and Spanish groups, and a weaker cluster in the Farsi and African groups. However, Vocabulary splits off from Reading Comprehension to form a clearly separate factor in the Chinese, Japanese, and Total groups and shows some tendency to split off from Reading Comprehension in the African, Arabic, Farsi, and Germanic groups.

Table 2

Number of Items with Highest Loading on
Each Factor/Four-factor Varimax Rotation

Language Groups		n	Factor I	II	III	IV
African	Listening Comp.	50	45*	2	2	1
	Structure	14(2)#	1	5	7*	1
	Written Expres.	25(1)	0	7	18*	0
	Vocabulary	30(1)	4	22*	3	1
	Reading Comp.	30(1)	3	10	12	5
	Variance		12.5	11.3	10.8	3.6
	Roots	29.8 3.9 3.5	2.6 2.1	2.0 2.0	1.8 1.7 1.6 1.6	
	Trace	53.38##				Σh^2 = 38.14
Arabic	Listening Comp.	50	44*	3	1	2
	Structure	14	5	6	1	2
	Written Expres.	25	1	9	4	11
	Vocabulary	30	2	16*	12	0
	Reading Comp.	30	1	21*	6	2
	Variance		16.8	14.7	7.9	6.7
	Roots	36.3 4.2 3.4	2.3 2.1	1.9 1.8	1.7 1.7 1.6	
	Trace	57.13				Σh^2 = 46.02
Chinese	Listening Comp.	50	8	37*	4	1
	Structure	14	8*	1	4	1
	Written Expres.	25	16*	1	6	2
	Vocabulary	30	7	0	23*	0
	Reading Comp.	30	26*	0	2	2
	Variance		17.2	13.1	8.6	3.0
	Roots	33.0 3.9 3.4	1.9 1.6	1.5 1.4	1.3	
	Trace	48.72				Σh^2 = 42.18
Farsi	Listening Comp.	50	44*	3	2	1
	Structure	14	5	3	4	2
	Written Expres.	25	3	5	5	12
	Vocabulary	30	3	10	12	5
	Reading Comp.	30	1	24*	3	2
	Variance		15.0	11.0	6.5	6.0
	Roots	29.3 4.1 3.5	1.8 1.6	1.5 1.4	1.3 1.3 1.2	
	Trace	47.74				Σh^2 = 38.42

(Cont'd.)

Table 2 (Cont'd.)

Language Groups		n	Factor I	II	III	IV
Germanic	Listening Comp.	50(6)	37*	13	0	0
	Structure	14(2)	6	3	5	0
	Written Expres.	25(3)	2	10	10	3
	Vocabulary	30(5)	9	7	12	1
	Reading Comp.	30(8)	1	23*	5	1
	Variance		12.5	12.3	9.0	3.0
	Roots 33.5 6.5 4.0		3.2 3.0 2.7 2.5			
	Trace 56.75					Σh^2 = 36.30
Japanese	Listening Comp.	50	42*	3	3	2
	Structure	14	1	7*	2	4
	Written Expres.	25	0	18*	2	5
	Vocabulary	30	3	0	23*	4
	Reading Comp.	30(1)	5	13	6	6
	Variance		13.6	11.2	9.6	5.3
	Roots 31.8 4.2 2.3		1.7 1.6 1.5 1.5 1.4			
	Trace 47.15					Σh^2 = 39.63
Spanish	Listening Comp.	50	45*	4	1	0
	Structure	14	7*	4	2	1
	Written Expres.	25	6	15*	1	3
	Vocabulary	30(6)	2	8	16*	4
	Reading Comp.	30(2)	1	2	25*	2
	Variance		30.3	13.3	12.2	4.8
	Roots 43.6 4.7 3.3		1.9 1.7 1.6 1.5 1.4			
	Trace 60.49					Σh^2 = 50.50
Total	Listening Comp.	50	7	34*	6	3
	Structure	14	3	1	6	3
	Written Expres.	25	19*	0	5	1
	Vocabulary	30	4	3	19*	4
	Reading Comp.	30	26*	1	3	0
	Variance		16.5	16.1	1.3	7.6
	Roots 42.7 4.8 3.0		2.6 1.7 1.6 1.4 1.4			
	Trace 59.73					Σh^2 = 53.00

*Asterisks represent cases in which half or more of the items of a given type have their highest loadings on the indicated factor.

#Numbers of items in parentheses were passed by 96 percent or more of subjects. They were not used to determine factor structure, but loadings for them were obtained through extension analysis.

##The trace given here is 149 X (the average Tucker communality estimate). Thus Σh^2/Trace estimates the proportion of common variance accounted for by the four factors.

Thus, although Listening Comprehension items form a separate dimension in all groups, Written Expression clusters with Structure for some native languages, but does not for others, and Vocabulary forms a separate dimension from Reading Comprehension in most groups. In the cases in which Vocabulary forms the most unequivocally independent factor (African, Chinese, and Japanese), Structure, Written Expression, and Reading Comprehension tend to load on a single factor. In no case do both Structure and Vocabulary items display half or more of their loadings on the same factor. This last observation allows us to perform further Procrustes rotations on these minres factor solutions to maximize comparability without distorting the empirical factor structure in any group.

Rotation to an A Priori Target

The four-factor minres solutions for each language group were subjected to an orthogonal Procrustes rotation (Browne & Kristof, 1969) to force them to fit the design structure of the test, to the extent that the data would allow, and to facilitate comparisons among the subject groups. The target structure consisted of three constrained factors and a fourth unconstrained factor. The first target factor was specified as maximum loadings on 14 items of middle difficulty from the Listening Comprehension section of the test, with loadings on other items unspecified.

The second target factor was specified as maximum loadings on all 14 Structure items from Section II of the test, with loadings on other items unspecified. The third target factor consisted of 14 Vocabulary items of middle difficulty from Section III of the test.

Our rationale for performing these rotations was, first, that we wished to facilitate comparisons by orienting the factors of each group as closely as possible, within the constraints of orthogonality, to the specifications of the test, and, second, that we wished to allow departures from the specified structure to reveal themselves uniquely for each group. The number of items determining a target factor was limited to 14 because the structure section of this test form contained only 14 scored items. If more than 14 items from another section had been specified as the target for another factor, the rotation would have tended to obtain a spuriously better fit to the factor having more target items.

If all Listening Comprehension items do indeed form a single factor, the remaining nontarget 36 listening items would be expected to load on Factor I just as strongly as the target items do. The target items themselves should, of course, all load more strongly on Factor I than on the other three factors. On the other hand, if the Listening Comprehension section taps more than one dimension of performance, no orthogonal rotation will force the set of items to fit a single factor well. Similarly, if Structure and Written Expression items form a single dimension, the Written Expression items should coalesce with the Structure

items to fit target Factor II. That Written Expression items were not specified as part of the target matrix, however, makes it possible for this expectation to be disconfirmed by the data. Although a weighting procedure could have been employed (e.g. weighting Listening Comprehension items each by 14/50), this would not have made possible the internal test of homogeneity afforded by assigning a weight of 1 to 14 items and a weight of "zero" to the remaining 36 Listening Comprehension items.

The same rationale applies to Section III of the test. Only 14 Vocabulary items are specified in the target. If performance on Section III is in fact unidimensional, the other 16 nontarget Vocabulary items and all 30 Reading Comprehension items should load on Factor III. The unspecified fourth factor allows room for items to split off from the specified structure if the data force such splitting. In fact, the fourth factor did not prove necessary in describing the structure of the test for any language group except the Germanic.

Interpretation of Target Factors

The results of the four-factor orthogonal rotations to target for each language group are shown in Appendix A, in which all loadings greater than .20 are given. Three summaries of those data will be given here:

1. the number of various TOEFL item types having the highest loading on each factor for each language group (Table 3);

2. the number of various TOEFL item types having high loadings (.30 or higher) on each factor for each language group (Table 4); and

3. a description of the particular individual item that had the highest loading of any item on each factor for each language group (Table 5).

Although each summary in Tables 3-5 is important, each is given in the order of importance to the interpretation of factors. That is, Table 3 deserves more attention than either Table 4 or Table 5, and so on.

Factor I, targeted on medium difficulty Listening Comprehension items, is clearly defined for each language group by a preponderance of high loadings from Listening Comprehension items. Depending on the particular language group, from 76 percent to 88 percent of Listening Comprehension items loaded higher on Factor I than on any other. In addition, from 76 percent to 90 percent of all Listening Comprehension items had loadings greater than .30. Finally, for each language group, the item having the highest loading appeared in the Listening Comprehension section of the test. Factor I, therefore, is labeled Listening Comprehension. Factor II, targeted on all Structure items, is less easily characterized than Factor I because of high loadings from several item types and because of differences from group to group. Factor II is most similar for the African, Arabic, and Japanese groups: all or nearly all Written Expression items and a majority of both Structure and Reading Comprehension

Table 3

Numbers of Various Types of TOEFL Items Loading Highest on Each of Four Factors for Seven Language Groups/Target Factors

Language Group	Item Type	Number of Items	Factors I	II	III	IV
	Listening Comprehension	50	44	4	1	1
	Structure	14	0	9	5	0
African	Written Expression	25	0	25	0	0
	Vocabulary	30	1	7	20	2
	Reading Comprehension	30	1	22	6	1
	Listening Comprehension	50	38	12	0	0
	Structure	14	5	8	1	0
Arabic	Written Expression	25	0	25	0	0
	Vocabulary	30	2	9	17	2
	Reading Comprehension	30	1	20	8	1
	Listening Comprehension	50	42	6	1	1
	Structure	14	1	10	3	0
Chinese	Written Expression	25	2	15	8	0
	Vocabulary	30	0	8	22	0
	Reading Comprehension	30	0	27	2	1
	Listening Comprehension	50	39	10	0	1
	Structure	14	4	5	5	0
Farsi	Written Expression	25	6	8	10	1
	Vocabulary	30	5	11	13	1
	Reading Comprehension	30	8	1	20	1
	Listening Comprehension	50	43	6	1	0
	Structure	14	1	8	4	1
Germanic	Written Expression	25	1	16	8	0
	Vocabulary	30	1	11	15	3
	Reading Comprehension	30	4	4	18	4
	Listening Comprehension	50	43	5	2	0
	Structure	14	0	10	1	3
Japanese	Written Expression	25	1	21	2	1
	Vocabulary	30	2	4	23	1
	Reading Comprehension	30	3	22	5	0
	Listening Comprehension	50	38	12	0	0
	Structure	14	3	7	4	0
Spanish	Written Expression	25	1	21	3	0
	Vocabulary	30	1	9	19	1
	Reading Comprehension	30	1	2	27	0

Table 4

Numbers of Various Types of TOEFL Items Having Loadings Greater than .30 on Each of Four Factors for Seven Language Groups/Target Factors

Language Group	Item Type	Number of Items	Factors I	II	III	IV
	Listening Comprehension	50	38	4	1	0
	Structure	14	2	9	5	1
African	Written Expression	25	0	21	2	1
	Vocabulary	30	2	7	17	0
	Reading Comprehension	30	4	17	5	6
	Listening Comprehension	50	42	26	2	0
	Structure	14	6	6	3	1
Arabic	Written Expression	25	4	24	5	2
	Vocabulary	30	4	8	19	3
	Reading Comprehension	30	4	21	16	2
	Listening Comprehension	50	39	12	0	0
	Structure	14	3	8	5	0
Chinese	Written Expression	25	3	14	4	3
	Vocabulary	30	1	11	17	0
	Reading Comprehension	30	1	25	4	3
	Listening Comprehension	50	40	24	3	1
	Structure	14	6	5	3	1
Farsi	Written Expression	25	6	10	7	2
	Vocabulary	30	1	9	11	0
	Reading Comprehension	30	14	3	20	0
	Listening Comprehension	50	42	13	5	2
	Structure	14	2	10	4	3
Germanic	Written Expression	25	2	18	11	1
	Vocabulary	30	1	14	10	3
	Reading Comprehension	30	12	13	20	9
	Listening Comprehension	50	40	8	4	1
	Structure	14	3	10	2	1
Japanese	Written Expression	25	1	19	4	2
	Vocabulary	30	2	5	17	0
	Reading Comprehension	30	7	24	17	0
	Listening Comprehension	50	45	32	7	0
	Structure	14	6	9	5	0
Spanish	Written Expression	25	6	21	5	1
	Vocabulary	30	6	9	15	2
	Reading Comprehension	30	12	8	28	0

Table 5

Characteristics of Item Loading Highest on
Each Factor for Each Language Group/Target Factors

Factor	Language Group	Size of Loading	Item Number
I	African	.56	LC-11
	Arabic	.68	LC-31
	Chinese	.61	LC-32
	Farsi	.68	LC-27
	Germanic	.66	LC-48
	Japanese	.63	LC-25
	Spanish	.68	LC-27
II	African	.82	WE-29
	Arabic	.64	WE-16,17,33
	Chinese	.64	RC-52
	Farsi	.51	S-10
	Germanic	.80	WE-22
	Japanese	.59	RC-44
	Spanish	.62	V-18
III	African	.67	V-30
	Arabic	.60	RC-52
	Chinese	.58	V-28
	Farsi	.55	RC-59
	Germanic	.71	RC-59
	Japanese	.60	V-25
	Spanish	.59	V-25,RC-5
IV	African	.44	WE-16
	Arabic	.25	RC-46
	Chinese	.55	RC-35
	Farsi	.36	S-8
	Germanic	.50	V-4
	Japanese	.40	S-8
	Spanish	.40	WE-16

items loaded higher on this factor than on any other. The Chinese group also had a similar pattern of loadings, but with a smaller majority (60 percent) of Written Expression items loading highest on this factor. The pattern of loadings greater than .30 is also similar for this factor among the groups mentioned. In addition, the items loading highest on this factor for both the African and Arabic groups are Written Expression items, but are Reading Comprehension items for the Chinese and Japanese groups. Thus, for the African, Arabic, Chinese, and Japanese groups, Factor II seems to underlie three different item types: two dealing with writing (Structure and Written Expression) and one dealing with reading (Reading Comprehension).

A pattern similar to that described above was noted for Factor II for the Spanish and Germanic groups, with one important difference: Reading Comprehension items did not load particularly strongly on Factor II, but instead tended to load on Factor III, the Vocabulary target factor, thus fitting the specifications of the three-part TOEFL well. For the Farsi group, Reading Comprehension items also loaded mainly on Factor III, instead of on Factor II, and neither Structure nor Written Expression items loaded as strongly on Factor II as these items did for other language groups. These three language groups constitute the Indo-European languages in this sample.

Factor III is defined and targeted by high loadings from Vocabulary items for the African, Arabic, Chinese, and Japanese groups. For the Germanic and Spanish groups, Factor III is also defined by high loadings from a majority of Vocabulary items, but even more strongly characterized by Reading Comprehension items, which were carried along in the rotation to the vocabulary target. For the Farsi group, Factor III is defined by both Reading Comprehension and Vocabulary items and also to a lesser extent by written expression items.

Factor IV is not readily linked to any particular item type(s) for most language groups. This "residual factor" was carried in the analysis in order to give any interpretable orthogonal residual variance the chance to emerge as a factor. For the African language group, however, Factor IV is quite clearly interpretable as reading speed, while for the Germanic group, this factor contrasts Reading Comprehension and Vocabulary items. A summary of factors for the individual varimax solutions and the solutions rotated to target is given in Table 6.

Discussion

Integrating Findings with Other Research

The present study offers some evidence that (a) three major factors underlie performance on the TOEFL and (b) these factors are relatively unambiguous in their interpretation. A factor underlying the Listening

Table 6

Summary of Factors for Several Language Groups
for Two Methods of Analysis

	Method of Analysis							
	Individual Varimax Solutions				Individual Solutions Rotated to a Target			
Language Group	Factor I	Factor II	Factor III	Factor IV	Factor I	Factor II	Factor III	Factor IV
African	LC	V	WE,S,RC	--	LC	WE,RC,S	V	--
Arabic	LC	RC,V,S	--	WE	LC	WE,RC,S	V	--
Chinese	RC,WE,S	LC	V	--	LC	RC,S,WE	V	--
Farsi	LC	RC	V	WE	LC	(S)	RC,V,WE,(S)	--
Germanic	LC	RC,(WE)	V,(WE)	--	LC	WE,S	RC,V	--
Japanese	LC	WE,S,RC	V	--	LC	WE,RC,S	V	--
Spanish	LC	WE	RC,V	--	LC	WE,S	RC,V	--

KEY: LC = Listening Comprehension
S = English Structure
WE = Written Expression
V = Vocabulary
RC = Reading Comprehension

Note. Each item type is listed under the factor on which the item type has its greatest number of highest loadings, in order of the types having the highest percentage of highest loadings.

Types in parentheses indicate that the item type also had an equal number of highest loadings on another factor.

Comprehension section was noted for each language group. However, there are differences among the language groups in the interpretation of two of the factors. For the two Indo-European language groups whose languages are most similar to English, e.g., the Germanic and the Spanish groups, the second and third factors correspond with the TOEFL subscores (Structure + Written Expression and Reading Comprehension + Vocabulary) that are now reported. For most other groups (African, Arabic, Chinese, and Japanese) the Reading Comprehension items tend to load on the same factor with Structure and Written Expression, with Vocabulary splitting off from Reading Comprehension to define a third factor. For Farsi speakers, who in this sample are the least proficient of any group with respect to total TOEFL scores, the factor structure is the least differentiated of any group structure. With the exception of the Listening Comprehension factor, each of the other factors is less clearly interpretable. Items from each section are more likely to load on several factors, and each factor is more likely to have high loadings from several item types for the Farsi group than for any other language group. Thus the Farsi results are more suggestive of a single Listening Comprehension factor and a global factor underlying performance on Sections II and III than are the structures of the other groups.

It is particularly interesting to mention here the results of a study by Oller (1976) that was designed to test the extent to which language proficiency is divisible into separately assessable components. Oller (1979) reports that a factor analysis of TOEFL data for 159 Iranian adults in Tehran supported the indivisibility hypothesis of a single global factor. Another relevant study, conducted by Hinofotis (1976) and reported by Oller and Hinofotis (1976), in which data were obtained at the Center for English as a Second Language, was somewhat more supportive of the possibility of separate language skills, particularly speaking skills. The point to be made here is that the results of correlational studies of language skills may be quite dependent on the particular sample used, especially with respect to their overall level of language proficiency. One hypothesis that could be investigated is the extent to which separate factors (or components of variation) are more likely to emerge as the overall language proficiency of the sample increases. Information on this question might add to the understanding of the global proficiency vs. discrete-point language testing controversy. There is some evidence to support this hypothesis, particularly when one contrasts the factor structure of the TOEFL for Farsi speakers with the considerably more complex structure for Germanic speakers. The factor-analytic literature on mental abilities contains a number of studies suggesting that mental abilities are more highly differentiated in samples that are older or that have had more exposure to relevant instruction and experience.

Pike (1979) in his important monograph, "An Evaluation of Alternative Item Formats for Testing English as a Foreign Language," contrasted intercorrelations among the five item types that constituted the then five-part TOEFL, for Peruvian, Chilean, Japanese, and all TOEFL candidates. Table 7, after Pike's Table 7, gives the median correlations

Table 7

Correlations Among TOEFL Sections as Reported by Pike

	Subtest				
	Listening Comp.	English Structure	Written Expression	Vocab.	Reading Comp.
		Hispanic (n=243)			
Listening Comp.	[.94]	(.93)	(.87)	(.84)	(.81)
English Structure	.85	[.91]	(.97)	(.85)	(.82)
Written Expression	.79	.86	[.88]	(.88)	(.90)
Vocabulary	.74	.74	.76	[.85]	(.93)
Reading Comprehension	.71	.71	.71	.78	[.83]
		Japanese (n=199)			
Listening Comp.	[.85]	(.80)	(.65)	(.61)	(.74)
English Structure	.65	[.78]	(.82)	(.78)	(.85)
Written Expression	.52	.63	[.76]	(.84)	(.84)
Vocabulary	.51	.62	.66	[.82]	(.88)
Reading Comprehension	.56	.62	.61	.66	[.69]
		World (n=1,000)			
Listening Comp.	[.88]	(.74)	(.67)	(.63)	(.77)
English Structure	.63	[.83]	(.93)	(.82)	(.83)
Written Expression	.58	.77	[.82]	(.81)	(.83)
Vocabulary	.54	.68	.67	[.83]	(.84)
Reading Comprehension	.64	.67	.67	.68	[.79]

[ ] reliabilities

Note. Correlations above the diagonal have been corrected for attenuation.

for the two Hispanic groups, for the Japanese group, and for the "World" group (a sample of all candidates).

The underlined corrected correlations, r_1, between Structure and Written Expression, and r_2, between Vocabulary and Reading Comprehension, formed the basis for a conjecture and a recommendation in the Pike report. The conjecture, based on corrected correlations $r_1 = .97$ and $r_2 = .93$ for Hispanics, vs. $r_1 = .82$ and $r_2 = .88$ for Japanese speakers, was that "the relationships among component English language skills may tend to be lower for Japanese students than for those having an Indo-European language background" (p. 55). The recommendation was that since Structure and Written Expression form one cluster ($r_1 = .93$ for World) and the Reading Comprehension and Vocabulary measures form another ($r_2 = .84$) that "these and other considerations suggest a revised TOEFL having several components, but yielding only three scores: I. Listening Comprehension, II. English Structure and Writing Ability, and III. Reading Comprehension and Vocabulary in Context."

The results of the present factor analysis offer strong support to the conjecture, since none of the non-Indo-European language groups examined exhibited the clustering of Vocabulary and Reading Comprehension found in the Spanish and Germanic groups. The recommendation is supported in the case of the English Structure and the Written Expression measures, but considerably less strongly supported in the case of Reading Comprehension and Vocabulary. For non-Indo-European speakers, Reading Comprehension appears to be more strongly related to Structure and Written Expression than to Vocabulary. Indeed, the "World" corrected correlations presented in the Pike report for Vocabulary (.81 with Written Expression, .82 with Structure, and .84 with Reading Comprehension) do not offer convincing evidence that Vocabulary is significantly more related to Reading Comprehension than to the other two measures. Although the corrected correlations within the two clusters exceed .90 in the Hispanic groups, this is the case only for the Structure and Written Expression cluster for the "World" group, and for neither cluster for the Japanese group in the Pike data.

If analysis of TOEFL forms more recent than form YTF4 confirms that Reading Comprehension items indeed covary with Structure and Written Expression more strongly than with Vocabulary in non-Indo-European groups, an important consequence would emerge in interpreting TOEFL scores for members of such groups: In the case of discrepancies between Structure and Written Expression (Section II) total scores and Reading Comprehension & Vocabulary (Section III) total scores, non-Indo-European candidates are probably more discrepant in vocabulary than their Section III scores indicate, since half of this Section III score is based on Reading Comprehension, a skill which tends to pull the combined score toward the Structure and Written Expression score more than it does for

Indo-European speakers. This structural difference does not affect total TOEFL scores, but it has considerable implications for the interpretation of section scores among different groups. The difference in the relationship of vocabulary to other measures in the two broad language groups may be partly due to the different store of cognate and false cognate words among Indo-European vs. non-Indo-European languages, but it is probably due more to structural and syntactic similarities and differences. In the absence of evidence to the contrary, we concur with Fillmore (1968) that human semantic, more than syntactic, categories share elements of universality. The structural differences noted here should thus be taken into account in interpreting TOEFL scores, but not overinterpreted. A relatively high Section III score for a non-European candidate probably represents unusually high vocabulary achievement. A relatively low Section III score, when compared with the Section II score, probably reflects low attainment in vocabulary relative to reading comprehension ability. However, vocabulary in a second language is acquired more rapidly than is syntactic competence, and is thus more easily remediated for these candidates if grammatical knowledge, as measured by Section II, and Reading Comprehension, as measured by Section III, in are adequate. These considerations suggest differential interpretations of TOEFL scores for different language groups. They also support the position that a total TOEFL score may be less useful for assessing of language competence than the three section scores are, if interpreted in the light of the language background of the students. Further, if these findings are confirmed in analyses of other forms of TOEFL, the potential usefulness of a separate vocabulary score should be considered.

Tables 8 and 9 show for samples of Foreign candidates, (a) the intercorrelations among item types and sections on TOEFL form YTF4 and (b) the average correlations for six more recently developed TOEFL forms. In both cases, the correlation between Sections I and II is .70, as is that between Sections I and III. However, the correlation between Sections II and III drops slightly from .83 in Form YTF4 to an average .78 in the Spring 1980 3CTF series. In the form analyzed in this study, (YTF4) the disattenuated correlation between II and III is .94, those between I and II and between I and III are .80 and .79, respectively. Thus, in Form YTF4, Section III is so strongly related to Section II that it impairs its usefulness as a separate score. In Spring 1980 the corrected correlations between I and II was .80, and between I and III, .77. Thus they are comparable to those of Form YTF4; however, the average corrected correlation between II and III has dropped slightly to .89. When Spring 1980 item-type subscore correlations are corrected for attenuation, Reading Comprehension relates almost as strongly to Structure ($\underline{r}$ = .83), and to Written Expression ($\underline{r}$ = .84), as to Vocabulary ($\underline{r}$ = .86), although it is combined only with Vocabulary scores to create Section III scores. However, the corrected correlations of Vocabulary with Structure (.85) and with Written Expression (.83) are comparable to the corrected correlation of .86 with Reading Comprehension. The corrected correlation between Structure and Written Expression is .89, very slightly greater than that for other pairs of item types. Therefore, both the factor-

Table 8

Section Intercorrelations n=1135

FORM YTF4 November 1976

	Listening Comp.	Structure	Written Expression	II Structure and Written Expression	Vocabulary	Reading Comp.	III Vocabulary and Reading Comp.
I = Listening Comp.	.90	.64	.66	.70	.61	.70	.71
Structure		.71	.69	(.87)	.67	.69	.73
Written Expression			.82	(.96)	.70	.77	.79
II = Structure and Written Expression				.87	.74	.80	.83
Vocabulary					.83	.72	(.92)
Reading Comp.						.88	(.93)
III = Vocabulary and Reading Comp.							.92

Note. Reliabilities in main diagonal (part-whole correlations in parenthesis)

Table 9

Average Section Intercorrelations, Six TOEFL Forms
February - July 1980

	Listening Comp.	Structure	Written Expression	II Structure and Written Expression	Vocabulary	Reading Comp.	III Vocabulary and Reading Comp.
I = Listening Comp.	.894	.626	.649	.698	.623	.662	.687
Structure		.709	.662	(.858)	.653	.630	.692
Written Expression			.784	(.953)	.671	.671	.725
II = Structure and Written Expression				.856	.722	.714	.777
Vocabulary					.842	.710	(.927)
Reading Comp.						.806	(.921)
III = Vocabulary and Reading Comp.							.899

Note. Reliabilities in main diagonal (part-whole correlations in parentheses).

analytic results and the operational-section statistics seem to concur that the marriage of TOEFL Vocabulary and Reading Comprehension scores to form Section III may not have been "made in heaven." If separate subscores are deemed to be important, it would appear that for non-Indo-European native speakers, Section III results might be more usefully reported as Vocabulary and Reading subscores than as a combined score. The slight trend over time toward decrease in correlations of Sections II and III apparently results from improved test specifications since the development of Form YTF4, but the trend would have to continue if the combined Section III scores are to become above psychometric reproach.

Factor Extension Analysis of Background Variables

Profiles of native language groups on background variables reveal similarities and differences pertinent to comparing factor structures across groups. Each of the seven groups in the study is predominantly male, although African and Arabic groups are distinctly so, with 87 percent and 89 percent of their respective examinees being male. Groups seem to differ more, however, with respect to the reasons for taking TOEFL. For all groups, the reason most often cited is support of undergraduate or graduate admissions. The proportion of examinees citing other choices (e.g., professional licensure or employment requirement) varies across groups. Germanic, Farsi, Spanish, and Japanese, for example, are similar in containing fewer undergraduates than graduates, roughly in a ratio of 3 to 5 in each group. Chinese and African, on the other hand, contain more undergraduates than graduates, in approximate ratios of 3 to 1 and 8 to 1, respectively. In addition, three groups, Arabic (21 percent), Japanese (17 percent), and Chinese (17 percent) contain a notable proportion of examinees taking TOEFL for reasons other than undergraduate or graduate admissions, such as professional licensure, business requirements, or admission to some other school program. A third background variable, previous TOEFL administrations, also reveals some marked contrasts among language groups. Only the Germanic and African groups have fewer than ten percent of the examinees with previous TOEFL experience. In other groups, the percentages range from 14 (for Spanish) to 42 (for Japanese). Finally, all groups are similar in having a majority of examinees destined for degree programs, although Germanic and Japanese groups have smaller percentages in this category than do the others. These data are summarized in Table 10.

Those variables most clear in reflecting group differences are "Reasons for Taking TOEFL"--largely a measure of graduate/undergraduate status and confounded with age--and "Previous TOEFL Administrations." The Chinese and African groups differ from others on the first of these variables, and the Japanese, African, and Germanic groups differ on the second.

Table 10 also shows the locations of factor extension correlations of at least $\pm$ .20 in magnitude and provides a rough idea of those

Table 10

Locations of Extension Correlations Greater Than or Equal to .20

Variable	Factor I	II	III	IV	Language Group	
Sex					German	(-)
					Spanish	(-)
					Farsi	(0)
					African	(+)
					Chinese	(-)
			-.27		Japanese	(0)
					Arabic	(+)
					TOTAL	
Birth Year					German	
		.22	-.21		Spanish	
					Farsi	
			-.31		African	
					Chinese	
			-.25		Japanese	
	.21			-.27	Arabic	
			-.24		TOTAL	
Undergraduate					German	(0)
		.26	-.26		Spanish	(0)
		.27	-.21		Farsi	(0)
			-.23		African	(+)
					Chinese	(0)
			-.21		Japanese	(0)
				-.32	Arabic	(0)
					TOTAL	
Graduate					German	(0)
			.20		Spanish	(0)
		-.27	.23		Farsi	(0)
			.27		African	(-)
					Chinese	(0)
			.21		Japanese	(0)
				.22	Arabic	(0)
			.26		TOTAL	
Studying for Degree					German	(-)
					Spanish	(+)
					Farsi	(+)
					African	(+)
					Chinese	(+)
					Japanese	(-)
					Arabic	(0)
					TOTAL	
Previous TOEFL					German	(-)
					Spanish	(0)
					Farsi	(0)
					African	(-)
					Chinese	(0)
					Japanese	(+)
					Arabic	(0)
				.22	TOTAL	

Key: - = Group low on this variable
+ = Group high on this variable
0 = Group average on this variable

background variables that may influence the interpretation of factor structures across groups. Here the most obvious trend is the occurrence of extension correlations of specified magnitude on variables related to age and/or graduate/undergraduate status. Such correlations occur consistently on Factor III, regardless of group profiles on those variables, and suggest a rather uniform relationship between the combination of age and academic status and Factor III. In particular, age and graduate are positively correlated with Factor III, suggesting that vocabulary develops with experience/exposure.

The remainder of Table 10 displays much of variation in the way that background variables correlate with factors common to the seven language groups. For the age/academic status combination mentioned above, Spanish, Farsi, African, and Arabic groups have extension correlations on at least two of the three other factors. Sex appears to be another variable that correlates differentially across groups, ; with females, e.g., showing superior knowledge of vocabulary, especially in the Japanese group. Although these are precisely the kinds of differences that may account for contrasting factor structures, it is not clear that the clustering of groups is the same pattern of clustering demonstrated by the extension correlations. That is, groups that are similar in profile on a given background variable like graduate status do not necessarily have a similar pattern of extension correlations. More detailed factor extension correlation matrices are presented in Appendix B.

Other Approaches

The present study is necessarily somewhat restricted in its approach. The restrictions apply mainly to the use of exploratory factor analysis with orthogonal rotations. Alternative confirmatory factor analytic approaches (e.g., Rock & Werts, 1979) could have been used to test the goodness of fit of TOEFL data to a factor structure specified a priori. Oblique rotations, in which factors are not constrained to be uncorrelated, would also have been appropriate, given that to require language proficiency factors to be totally independent gives descriptive economy at some cost in substantive interpretability. Item response theory might also have been applied to look at the relationships among component skills for language groups. Finally, the inclusion of marker variable tests (i.e., tests measuring several language skills in addition to those included in TOEFL), such as were included in some of the previous studies mentioned in Oller (1979), would have facilitated interpretations of the various factors. However, even within these limitations, the impressively consistent replications from all the included non-Indo-European language groups suggest that Vocabulary and Reading Comprehension as measured by TOEFL do not form a single dimension, a finding that has implications for score interpretation. In particular, when a non-Indo-European speaker exhibits Section III scores (Reading Comprehension and Vocabulary) discrepant from his or her Section II scores (Structure and Written Expression), the difference is likely to be due more to vocabulary than

to reading comprehension and to be greater than the score difference implies. This situation exists because part of the Section III score is based on Reading Comprehension, and in these language groups, is more strongly correlated with Section II, hence bringing the average Section III score closer to the Section II score than happens with scores earned by the Indo-Europeans. That non-Indo-European Section III score differentials can be identified more closely with vocabulary competency than with reading ability offers some useful diagnostic information for teachers of English as a second language. Vocabulary is relatively easily developed once grammar and syntax have reached an adequate stage of performance. Thus if a non-Indo-European exhibits relatively lower Section III scores on TOEFL, vocabulary development, rather than work on reading comprehension and vocabulary, may be indicated. For such students reading comprehension appears to be measured as much by Section II as by Section III. This would suggest that Japanese students, who rank sixth on Section II but fifth on Section III, might be stronger in vocabulary than their section scores suggest and might be more in need of Structure, Written Expression and Reading Comprehension exercises, West Africans, in spite of their ranking third on Section III, are likely to be in more relative need of vocabulary development. Spanish speakers, on the other hand, appear from their means and from the factor interpretation to be relatively more in need of explicit structure and written expression training, but less in need of a vocabulary or reading comprehension focus until their syntactic and grammatical competence catch up with vocabulary and reading. The vocabulary and reading of these Spanish speakers do relate in such a way that a single Section III score can be a useful summary of both skills.

Summary

The present factor-analytic study was undertaken to determine the explanatory constructs or component abilities underlying performance on the TOEFL and to determine the similarity of these component abilities for several major language groups. The results were intended to be useful for the interpretation of TOEFL scores and for future test development activities.

Procedures involved drawing samples of candidates from seven major language groups: African, Arabic, Chinese, Farsi, Germanic, Japanese, and Spanish. Correlations among all items on the TOEFL were computed for each language group and for the total sample of candidates. Separate, uncorrelated-factor solutions were obtained for each language group and for the total group of examinees. Because an infinite number of rotations of factors can be chosen that fit equally well the observed correlations among variables, a four-factor target matrix was specified, toward which the solution of each language group was rotated for best fit. Factor extension analysis was employed to assess the relationships of several

candidate background characteristics with factors underlying test performance. This technique was used to facilitate the interpretation of factors.

Three or four factors appeared to be necessary for each language group in order to explain the bulk of the variability in test-score performance on the TOEFL. There was some indication that the factor structure of the test might be more complex for Germanic speakers, for whom the test is generally relatively easy. Although the nature of the additional components of test variance for this group was not readily interpretable, the presence of these components suggests that English language proficiency may be more highly differentiated for Germanic speakers. On the other hand, for Farsi speakers, for whom the test is relatively difficult, the factor structure is least clear and suggests fewer highly differentiated factors.

Both similarities and differences in structures between language groups were noted. A listening comprehension factor could be clearly interpreted for each group. Other factors, however, differed from group to group. The African, Arabic, Chinese, and Japanese groups were generally similar on a factor underlying performance on Structure, Written Expression, and Reading Comprehension items, and on another separate factor underlying Vocabulary items. The Spanish and Germanic groups were also similar on each of two other factors, which correspond to the TOEFL subscores (Structure/Written Expression, and Reading Comprehension/Vocabulary) that are reported.

The Vocabulary factor exhibited positive correlations with age and degree-intentions in nearly every language group, thus suggesting that vocabulary is the most likely of any of the abilities to develop with training or experience. In all of the language groups, Vocabulary items were more likely than any other item type, except Listening Comprehension, to form a separate factor, a finding which suggests that including a separate vocabulary score might perhaps be justified.

In summary, there is evidence that three major factors underlie performance on the TOEFL but that these factors may differ in their interpretation for different major language groups. The results also suggest rather strongly that interpretations depend on the level of language skills in the group being considered.

The results also have implications for the interpretation of TOEFL subscores. For certain language groups, particularly the non-Indo-European groups, discrepancies between Section II (Structure and Written Expression) and Section III (Reading Comprehension and Vocabulary) scores are more likely to have resulted from knowledge (or lack of knowledge) of English vocabulary than from reading comprehension ability when compared with Germanic and Spanish speakers. Thus, a relatively low Section III score for non-Indo-European groups may be less critical for these groups, since vocabulary may be learned more readily than grammatical or syntactical structure.

Detailed interpretation of individual item loadings on the factors found in this study, and of their relation to specific vocabulary or syntactic features of each item in each language might be one way of controlling the structure of an English proficiency test to make that structure invariant across the native languages of the examinees. We believe that John Ciardi's recent remark about etymology also applies to this incredibly complex task. Ciardi said, "The main trouble with etymology is that the minimum requirement for doing it is omniscience."

References

Alderman, D. L. TOEFL item performance across seven language groups. Report in progress to the TOEFL Research Committee, 1980.

American Psychological Association. Standards for educational and psychological tests. Washington, D.C.: Author, 1974.

Angelis, P. J., Swinton, S. S., & Cowell, W. R. The performance of non-native speakers of English on TOEFL and verbal aptitude tests. (TOEFL Research Report No. 3.) Princeton, N.J.: Educational Testing Service, 1979.

Angoff, W. H., & Sharon, A. A comparison of scores earned on the Test of English as a Foreign Language by native American college students and foreign applicants to U.S. colleges. (ETS Research Bulletin 70-8.) Princeton, N.J.: Educational Testing Service, 1970.

Browne, M. W., & Kristof, W. On the oblique rotation of a factor matrix to a specified pattern. Psychometrika, 1969, 34(3), 375-394.

Clark, J. L. D. The performance of native speakers of English on the Test of English as a Foreign Language. (TOEFL Research Report No. 2). Princeton, N.J.: Educational Testing Service, 1977.

Dwyer, P. S. The determination of the factor loadings of a given test from the known loadings of other tests. Psychometrika, 1937, 2, 173-178.

Fillmore, C. J. The case for case. In E. Bach and R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart and Winston, 1968.

Harman, H. H. Modern Factor analysis (3rd rev. ed.). Chicago: The University of Chicago Press, 1976.

Hinofotis, F. A. B. An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language. Unpublished doctoral dissertation, Southern Illinois University, 1976 (Reported in Oller, J. W. Language tests at school: A pragmatic approach. London: Longman Group Limited, 1979.)

Oller, J. W. Language tests at school: A pragmatic approach. London: Longman Group Limited, 1979.

Oller, J. W. Evidence for a general language proficiency factor: An expectancy grammar. Die Neuren Sprachen, 1976, 2, 165-174. (Reported in Oller, J. W. Language tests at school: A pragmatic approach. London: Longman Group Limited, 1979.)

Oller, J. W., & Hinofotis, F. A. B. Two mutually exclusive hypotheses about second language proficiency: Factor analytical studies of a variety of language tests. Paper read at the Winter Meeting of the Linguistic Society of America, Philadelphia, Pa. (Reported in Oller, J. W. Language tests at school: A pragmatic approach. London: Longman Group Limited, 1976.)

Pike, L. W. An evaluation of alternative item formats for testing English as a foreign language (TOEFL Research Report No. 2). Princeton, N.J.: Educational Testing Service, 1979.

Powers, D. E., & Swinton, S. S. The factor structure of the GRE Aptitude Test: Implications for test development and assembly. (RM 76-2) Princeton, N.J.: Educational Testing Service, 1976.

Rock, D. A., & Werts, C. E. Construct validity of the SAT across populations--An empirical confirmatory study (College Board RDR 78-79, No. 5; and ETS RR-79-2). Princeton, N.J.: Educational Testing Service, 1979.

Appendix A

Four-Factor Solution Rotated to Target for Each Language Group

Table A-1

Rotated Factor Loadings
African Sample

	I	II	III	IV	Difficulty
I-1	.41*	.26	--	--	.88
I-2	.50*	--	--	--	.52
I-3	.40*	--	--	--	.75
I-4	.39*	--	--	--	.74
I-5	.25	.39*	--	--	.76
I-6	.38*	.25	--	--	.52
I-7	.33*	--	--	--	.69
I-8	.50*	--	--	--	.61
I-9	.29*	.29	.24	--	.74
I-10	.45*	--	--	--	.64
I-11	.56*	.21	--	--	.66
I-12	.38*	--	.27	--	.45
I-13	.48*	--	--	--	.85
I-14	.40*	.22	--	--	.57
I-15	.48*	--	--	--	.35
I-16	.29*	.21	--	--	.33
I-17	.35*	--	--	--	.53
I-18	.33*	--	--	--	.40
I-19	.40*	.28	.20	--	.59
I-20	.34*	.25	--	--	.34
I-21	*	--	--	--	.41
I-22	.41*	.36	--	--	.91
I-23	.49*	--	--	--	.72
I-24	.41*	.29	--	--	.62
I-25	.43*	--	--	--	.72
I-26	.41*	.27	--	--	.75
I-27	.53*	.22	--	--	.68
I-28	.25*	--	--	--	.67
I-29	.44*	--	.26	--	.33
I-30	.45*	.21	--	--	.58
I-31	.47*	--	.25	--	.33
I-32	.53*	--	--	--	.40
I-33	.47*	--	.24	--	.75
I-34	.42*	--	--	--	.52
I-35	.38*	--	.24	--	.57
I-36	.35*	--	.27	--	.65
I-37	--	--	--	*	.65
I-38	.25	.33*	.21	--	.83
I-39	.22	--	.34*	--	.18
I-40	.24	.27*	.25	--	.72
I-41	.29	.30*	.24	--	.55

(Cont'd.)

*Largest loading of items

Table A-1

Rotated Factor Loadings
African Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.29*	--	--	--	.63
I-43	.52*	.23	--	--	.44
I-44	.41*	.28	--	--	.55
I-45	.28*	.22	.27	-.26	.83
I-46	.31*	--	--	--	.32
I-47	.34*	--	--	--	.35
I-48	.35*	--	--	.22	.31
I-49	.49*	.27	--	--	.64
I-50	.32*	--	--	--	.75

*Largest loading of items

Table A-1

Rotated Factor Loadings
African Sample

	I	II	III	IV	Difficulty
II-1	.31	.33*	.22	--	.92
II-2	--	.29	.39*	--	.87
II-3	.25	.42*	--	-.42	.94
II-4	--	.34*	--	--	.96
II-5	--	.34*	.20	--	.83
II-6	--	--	.35*	--	.76
II-7	--	.45*	--	--	.95
II-8	--	.21*	--	--	.69
II-9	--	.32	.57*	--	.68
II-10	--	.45*	.23	--	.87
II-11	.27	.28	.34*	--	.67
II-12	--	.49*	--	--	.85
II-14	.33	.20	.35*	--	.40
II-15	--	.33*	--	--	.71
II-16	--	.64*	--	-.44	.97
II-17	--	.66*	--	-.24	.93
II-18	--	.29*	--	--	.80
II-19	--	.46*	.28	--	.72
II-20	.22	.27*	--	--	.65
II-21	--	.38*	--	--	.83
II-22	.20	.74*	--	-.24	.92
II-23	.25	.47*	--	--	.73
II-24	--	.43*	--	--	.68
II-25	--	.53*	.34	--	.61
II-26	.23	.42*	.42	--	.51
II-27	--	.58*	--	--	.84
II-28	--	.62*	--	--	.88
II-29	--	.82*	--	--	.91
II-30	--	.65*	--	--	.82
II-31	--	.32*	--	--	.52
II-32	--	.22*	--	--	.45
II-33	--	.67*	--	--	.82
II-34	.22	.31*	.30	--	.50
II-35	.24	.45*	.23	--	.61
II-36	--	.57*	.08	--	.72
II-37	--	.41*	--	--	.45
II-38	--	.29*	.25	--	.47
II-39	--	.36*	--	--	.56
II-40	.27	.35*	.23	--	.58

*Largest loading of items

Table A-1

Rotated Factor Loadings
African Sample

	I	II	III	IV	Difficulty
III-1	--	.23	.49*	--	.71
III-2	--	.45*	--	--	.97
III-3	--	--	.24*	--	.80
III-4	--	.29*	.28	--	.66
III-5	--	--	.41*	--	.68
III-6	--	.28	.43*	--	.67
III-7	.20*	--	--	--	.70
III-8	--	--	.44*	--	.77
III-9	.42	.27	.43*	--	.77
III-10	--	--	--	-.20*	.91
III-11	--	.26*	.26	--	.77
III-12	--	*	--	--	.69
III-13	--	.33*	--	--	.88
III-14	--	.33	.34*	--	.90
III-15	--	.26	.39*	--	.56
III-16	--	--	.29*	--	.77
III-17	--	--	--	*	.43
III-18	.33	.36	.37*	--	.77
III-19	.28	.30*	.25	--	.81
III-20	--	.21	.27*	--	.53
III-21	--	.24	.35*	--	.66
III-22	--	.24*	.21	--	.79
III-23	--	.21	.39*	--	.81
III-24	--	--	.62*	--	.51
III-25	--	.40	.61*	--	.52
III-26	--	.22	.45*	--	.75
III-27	--	.37	.46*	--	.64
III-28	.11	--	.45*	-.27	.47
III-29	.20	.22	.31*	--	.43
III-30	.27	.21	.67*	--	.24
III-31	--	.43*	.28	--	.80
III-32	--	.23*	.23	--	.95
III-33	.29*	--	.20	-.24	.45
III-34	.23	.23*	--	--	.80
III-35	.28	.32*	--	--	.85
III-36	.28	.29*	--	--	.63
III-37	--	.35*	.30	--	.83
III-38	.20	.27	.33*	--	.74
III-39	.28	.34*	.23	--	.69
III-40	--	--	.21*	--	.77
III-41	.21	.37*	.3	--	.74

(Cont'd.)

*Largest loading of items

Table A-1

Rotated Factor Loadings
African Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	--	--	.26*	--	.74
III-43	--	.35*	--	--	.88
III-44	.30	.43*	.22	-.22	.71
III-45	.25	.26	.35*	--	.63
III-46	.28	.48*	--	--	.90
III-47	.24	.5[illegible]*	--	--	.80
III-48	.21	.24*	.23	--	.45
III-49	--	.42*	--	.25	.81
III-50	--	--	*	--	.55
III-51	.30	.35*	--	.27	.46
III-52	.23	.39*	--	-.20	.50
III-53	--	--	--	.30*	.55
III-54	.29	.45*	--	.31	.74
III-55	.27	.49*	--	.32	.72
III-56	.26	.44*	--	.42	.65
III-57	.25	.25	.23	.29*	.38
III-58	.28	.44*	--	.32	.59
III-59	.30	--	.34*	.21	.41
III-60	.32	.35*	--	.34	.64
Variance	11.67	15.82	9.07	2.96	
Coefficients of Congruence	.981	.957	.906		

*Largest loading of items

Table A-2

Rotated Factor Loadings
Arabic Sample

	I	II	III	IV	Difficulty
I-1	.36	.39*	--	--	.90
I-2	.41*	.24	--	--	.61
I-3	.45*	.26	--	--	.70
I-4	.45*	.22	--	--	.81
I-5	.21	.44*	--	--	.80
I-6	.39	.39*	.20	--	.62
I-7	.42*	.28	--	--	.69
I-8	.41*	.30	--	--	.69
I-9	.23	.28*	--	--	.69
I-10	.44*	.36	.25	--	.53
I-11	.46*	.39	.28	--	.59
I-12	.53*	.25	.21	--	.62
I-13	.48*	.47	--	--	.87
I-14	.48*	.37	--	--	.64
I-15	.39	.41*	--	--	.55
I-16	.44*	--	.20	.25	.26
I-17	.50*	.27	.24	--	.57
I-18	--	.20*	--	--	.42
I-19	--	.27*	.20	--	.45
I-20	.26	.34*	.2	--	.35
I-21	.43*	--	--	--	.76
I-22	.46*	.46	--	-.20	.92
I-23	.44*	.38	--	--	.88
I-24	.37*	.34	--	--	.75
I-25	.50*	.22	--	--	.77
I-26	.49*	.36	--	--	.82
I-27	.56*	.36	--	--	.79
I-28	.41*	--	--	--	.76
I-29	.41*	.35	--	--	.64
I-30	.60*	.32	--	--	.73
I-31	.68*	.23	.24	--	.51
I-32	.51*	.28	--	--	.49
I-33	.45*	.32	--	--	.70
I-34	.52*	.30	--	--	.64
I-35	.36*	.33	--	--	.58
I-36	.43*	--	--	--	.74
I-37	--	.22*	--	--	.51
I-38	.36	.37*	.22	--	.77
I-39	.32*	--	.32	--	.22
I-40	.43*	.32	.23	--	.74
I-41	.23	.34*	.28	-.21	.57

*Largest loading of items

(Cont'd.)

Table A-2

Rotated Factor Loadings
Arabic Sample

	I	II	III	IV	Difficulty
I-42	.42*	.28	--	--	.70
I-43	.43*	.36	--	--	.53
I-44	.44*	.25	.21	--	.75
I-45	.50*	.27	.21	-.20	.82
I-46	.22	.24*	--	--	.38
I-47	.39*	.20	.32	--	.50
I-48	.50*	.29	.26	--	.45
I-49	.57*	.31	.28	--	.68
I-50	.46*	.38	.21	--	.64

*Largest loading of items

Table A-2

Rotated Factor Loadings
Arabic Sample

	I	II	III	IV	Difficulty
II-1	.38*	.31	.20	--	.80
II-2	.31*	.21	--	--	.72
II-3	.30*	.21	.29	-.26	.93
II-4	.22	.37*	.25	--	.81
II-5	.33	.36*	--	--	.77
II-6	--	.22*	.20	--	.67
II-7	--	.29*	--	--	.57
II-8	.33*	--	--	-.30	.46
II-9	--	.25*	--	--	.41
II-10	.39	.44*	.36	--	.53
II-11	.26*	.26	.23	--	.54
II-12	--	--	.36*	--	.36
II-14	--	.44*	.29	--	.17
II-15	.27	.44*	.39	--	.45
II-16	.40	.64*	--	--	.94
II-17	.36	.64*	--	-.31	.90
II-18	--	.39*	--	--	.65
II-19	--	.34*	.26	--	.58
II-20	--	.37*	--	--	.78
II-21	.32	.50*	--	--	.76
II-22	.32	.52*	.26	-.20	.80
II-23	--	.52*	--	--	.73
II-24	--	.42*	.27	--	.56
II-25	--	.45*	.25	--	.64
II-26	--	.50*	.36	--	.50
II-27	--	.54*	--	-.20	.76
II-28	--	.48*	--	--	.76
II-29	.22	.52*	.30	-.30	.75
II-30	.22	.32*	.31	--	.52
II-31	.24	.36*	--	--	.50
II-32	--	.44*	--	-.25	.32
II-33	--	.64*	--	--	.66
II-34	--	.34*	--	--	.39
II-35	--	.37*	.36	--	.49
II-36	.25	.40*	.20	--	.38
II-37	--	.57*	.25	--	.42
II-38	--	.32*	.30	.21	.29
II-39	--	*	--	--	.26
II-40	--	.42*	.21	--	.26

*Largest loading of items

Table A-2

Rotated Factor Loadings
Arabic Sample

	I	II	III	IV	Difficulty
III-1	--	.28*	.23	--	.83
III-2	.25	--	.42*	--	.85
III-3	.33	.27	.45*	-.27	.70
III-4	.31	.35	.48*	--	.66
III-5	--	--	.32*	--	.53
III-6	.26	.38*	--	-.27	.92
III-7	.27	--	.35*	--	.81
III-8	--	.26	.38*	--	.63
III-9	.48*	.32	.32	--	.71
III-10	--	--	.22	.26*	.67
III-11	--	*	--	--	.69
III-12	.39*	--	--	--	.70
III-13	--	--	.43*	--	.56
III-14	--	.26	.45*	-.24	.50
III-15	--	--	--	.32*	.36
III-16	--	.22	.46*	--	.28
III-17	--	--	.23*	--	.46
III-18	.25	.39*	.32	--	.75
III-19	.24	--	.32*	.20	.47
III-20	--	--	.28*	--	.49
III-21	.23	--	.39*	--	.32
III-22	--	.26	.35*	.33	.41
III-23	--	--	.35*	--	.33
III-24	--	.37*	.32	--	.35
III-25	--	.45*	.28	--	.37
III-26	--	.43*	.31	.22	.30
III-27	.21	--	.26*	--	.30
III-28	--	--	.44*	.38	.22
III-29	--	.45*	.37	--	.24
III-30	--	*	--	--	1.13
III-31	.27	.42	.43*	--	.66
III-32	.34	.37*	--	--	.89
III-33	--	--	.32*	--	.42
III-34	.27	.28*	.26	-.24	.80
III-35	.35*	.29	.24	-.34	.85
III-36	.22	.28*	--	-.20	.65
III-37	.28	.33	.35*	-.21	.79
III-38	.30	.37*	.35	-.25	.64
III-39	--	.41*	.26	--	.74
III-40	--	.29*	.23	--	.70
III-41	.21	.47*	.29	-.29	.73

(Cont'd.)

*Largest loading of items

Table A-2

Rotated Factor Loadings
Arabic Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	.29	.32*	--	--	.77
III-43	.20	.36*	--	-.20	.76
III-44	.20	.51*	.24	--	.58
III-45	.21	.31*	.28	--	.51
III-46	.34	.39	.41	.41*	.76
III-47	--	.49*	.22	--	.65
III-48	--	.20*	--	--	.41
III-49	--	.39*	.34	-.20	.73
III-50	.20	.27	.35*	-.23	.64
III-51	--	.42*	.35	--	.50
III-52	--	.29	.60*	--	.43
III-53	--	.31	.31*	--	.63
III-54	--	.43	.44*	--	.68
III-55	.20	.45*	.33	--	.69
III-56	--	.46*	.40	--	.61
III-57	--	.27	.32*	.28	.38
III-58	--	.50*	.40	--	.41
III-59	--	.47*	.40	--	.52
III-60	--	.49*	.20	--	.60
Variance	13.87	18.17	10.37	3.63	
Coefficients of Correlation	.983	.947	.964		

*Largest loading of items

Table A-3

Rotated Factor Loadings
Chinese Sample

	I	II	III	IV	Difficulty
I-1	.56*	--	--	--	.92
I-2	.45*	--	--	--	.72
I-3	.53*	--	--	--	.70
I-4	.42*	--	--	--	.71
I-5	.25*	.22	.25	.25	.88
I-6	.32*	.30	--	--	.66
I-7	.31*	.29	--	--	.70
I-8	.52*	.20	--	--	.69
I-9	--	.21*	--	--	.79
I-10	.41*	.26	--	--	.48
I-11	.53*	.37	--	--	.63
I-12	.39*	.28	--	--	.57
I-13	.46*	.39	--	--	.84
I-14	.29*	.25	.26	--	.52
I-15	.38*	.21	.26	--	.50
I-16	.23*	--	--	-.24*	.20
I-17	.41*	.24	--	--	.61
I-18	.20*	--	--	-.24*	.38
I-19	.22*	--	--	--	.51
I-20	.22	.29*	--	--	.44
I-21	.38*	--	--	--	.59
I-22	.44*	.24	--	.20	.92
I-23	.56*	.20	--	--	.86
I-24	.44*	.25	--	--	.53
I-25	.49*	--	.21	--	.75
I-26	.40*	.37	--	--	.84
I-27	.48*	.29	--	--	.75
I-28	.48*	.27	--	--	.69
I-29	.45*	--	--	--	.72
I-30	.49*	.34	--	--	.66
I-31	.49*	.33	--	--	.48
I-32	.61*	.24	.24	--	.34
I-33	.36*	.23	--	-.23	.51
I-34	.47*	.28	--	--	.65
I-35	.30*	.29	--	--	.45
I-36	.39*	.25	--	--	.61
I-37	--	--	--	*	.60
I-38	.33	.48*	--	--	.84
I-39	.22*	--	--	-.20	.27
I-40	.34	.50*	--	--	.84
I-41	.27	.43*	--	--	.65

(Cont'd.)

*Largest loading of items

Table A-3

Rotated Factor Loadings
Chinese Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.23	--	.24*	--	.62
I-43	.43*	.32	--	--	.49
I-44	.57*	.38	--	--	.62
I-45	.34	.37*	--	--	.82
I-46	.35*	.25	--	--	.36
I-47	.33*	.29	.20	--	.58
I-48	.57*	.23	--	--	.42
I-49	.48*	.28	--	--	.66
I-50	.45*	.25	--	--	.70

*Largest loading of items

Table A-3

Rotated Factor Loadings
Chinese Sample

	I	II	III	IV	Difficulty
II-1	.22	.25	.31*	--	.73
II-2	.20	.23*	--	--	.70
II-3	.41	.44*	--	--	.90
II-4	.29	.30*	.28	--	.73
II-5	.26	.46*	--	--	.84
II-6	--	--	.25*	--	.64
II-7	.24	.24*	.20	--	.65
II-8	.33*	.26	--	--	.64
II-9	--	--	.36*	--	.52
II-10	.40	.46*	.42	--	.37
II-11	.25	.42*	.22	--	.55
II-12	.24	.43*	--	--	.30
II-14	--	.37*	.31	--	.41
II-15	--	.36*	.36	--	.39
II-16	.35	.54*	--	.26	.93
II-17	.40	.61*	--	.36	.91
II-18	.21	.47*	.20	--	.82
II-19	--	.40*	.32	--	.58
II-20	--	--	.21*	--	.76
II-21	--	.44*	.24	.35	.80
II-22	.22	.63*	--	.48	.88
II-23	--	--	.29*	--	.67
II-24	--	.38*	.21	--	.72
II-25	--	.43*	.20	.29	.81
II-26	--	.33*	.25	--	.51
II-27	.29	.53*	--	.23	.86
II-28	.23	--	.25*	--	.76
II-29	.31*	.39*	.27	--	.78
II-30	--	.38*	.32	--	.55
II-31	--	.20	.24*	--	.58
II-32	--	--	*	--	.29
II-33	--	.47*	.20	--	.82
II-34	--	.27	.29*	--	.63
II-35	--	.38*	.28	--	.73
II-36	.26	.31*	--	--	.49
II-37	.23	.24	.32*	--	.34
II-38	--	.22	.31*	--	.45
II-39	*	--	--	-.29	.27
II-40	.21	.29*	.25	-.20	.30

*Largest loading of items

Table A-3

Rotated Factor Loadings
Chinese Sample

	I	II	III	IV	Difficulty
III-1	--	.22	.41*	--	.66
III-2	--	.46*	.22	--	.89
III-3	.22	.42*	--	--	.76
III-4	.21	.29	.36*	--	.61
III-5	--	.25	.29*	--	.55
III-6	--	.36*	--	--	.72
III-7	--	--	*	--	.67
III-8	--	.21	.31*	--	.65
III-9	--	.28*	.26	--	.70
III-10	--	--	.38*	--	.48
III-11	--	--	.27*	--	.65
III-12	.27	.46*	--	.23	.79
III-13	--	.32	.34*	--	.67
III-14	--	.44	.46*	--	.36
III-15	--	.27	.35*	-.24	.20
III-16	--	.30*	.29	--	.53
III-17	--	--	*	--	.49
III-18	.31	.34*	.31	--	.55
III-19	--	--	.43*	--	.36
III-20	--	.37*	--	--	.51
III-21	.25	.26	.31*	--	.57
III-22	.27	.38	.46*	--	.37
III-23	--	--	.30*	--	.41
III-24	--	.25	.27*	--	.41
III-25	--	.24	.39*	--	.39
III-26	--	.30	.39*	--	.58
III-27	--	--	.23*	--	.46
III-28	--	--	.58*	-.27	.21
III-29	.21	--	.30*	-.27	.27
III-30	--	--	.38*	-.24	.13
III-31	.26	.46*	.25	--	.73
III-32	.22	.47*	--	--	.89
III-33	.20	.44*	--	--	.62
III-34	.26	.29	.36*	.34	.83
III-35	.50	--	.50	.55*	.92
III-36	--	.35*	.25	.24	.78
III-37	.23	.45*	.25	--	.82
III-38	.24	.33*	--	--	.68
III-39	--	.40*	--	--	.75
III-40	--	*	--	--	.74
III-41	.26	.50*	.21	--	.69

(Cont'd.)

*Largest loading of items

Table A-3

Rotated Factor Loadings
Chinese Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	--	.32*	--	.21	.70
III-43	.21	.36*	--	--	.86
III-44	.20	.56*	--	--	.66
III-45	--	.39*	.22	.23	.58
III-46	.27	.64*	--	--	.86
III-47	.29	.46*	.27	--	.67
III-48	--	.25*	.21	--	.47
III-49	.24	.47*	--	--	.84
III-50	--	.35*	--	--	.75
III-51	.21	.57*	--	--	.65
III-52	.21	.65*	.22	--	.50
III-53	--	.47*	--	--	.72
III-54	--	.51*	--	--	.85
III-55	.23	.61*	--	--	.81
III-56	.23	.55*	--	--	.68
III-57	--	.23	.37*	-.31	.40
III-58	.21	.64*	.26	--	.56
III-59	.21	.55*	.34	--	.59
III-60	.25	.45*	.24	--	.73
Variance	12.76	17.63	8.45	3.33	
Coefficients of Congruence	.984	.954	.969		

*Largest loading of items

Table A-4

Rotated Factor Loadings
Farsi Sample

	I	II	III	IV	Difficulty
I-1	.55*	.31	--	--	.84
I-2	.37*	.22	--	--	.52
I-3	.40*	--	--	--	.58
I-4	.44*	.24	--	--	.75
I-5	.48*	--	--	.25	.80
I-6	.40*	.32	--	--	.58
I-7	.32*	--	--	--	.63
I-8	.48*	.25	--	--	.65
I-9	.33*	--	--	--	.60
I-10	.27	.38*	--	--	.41
I-11	.51*	.39	--	--	.47
I-12	.42*	.37	--	--	.45
I-13	.37*	.27	--	--	.68
I-14	.35*	.34	--	--	.50
I-15	.47*	.39	--	--	.53
I-16	.23	.36*	--	.20	.32
I-17	.30	.33*	--	.22	.38
I-18	--	.26*	--	--	.41
I-19	--	--	--	.27*	.41
I-20	.21	.33*	--	--	.32
I-21	.27*	--	--	--	.73
I-22	.52*	.23	--	--	.89
I-23	.56*	.24	--	--	.74
I-24	.44*	.32	--	--	.59
I-25	.55*	.37	--	--	.73
I-26	.47*	.30	--	--	.68
I-27	.68*	.25	--	--	.73
I-28	.50*	.33	--	--	.57
I-29	.54*	.43	--	--	.55
I-30	.51*	.44	--	--	.55
I-31	.53*	.46	--	--	.35
I-32	.60*	.37	--	--	.37
I-33	.44*	.32	--	--	.44
I-34	.39*	.29	--	--	.55
I-35	.32	.37*	--	--	.35
I-36	.48*	--	.23	--	.58
I-37	--	*	--	--	.51
I-38	.46*	--	.31	--	.67
I-39	.28*	.24	.21	.21	.19
I-40	.44*	--	.25	--	.69
I-41	.40*	--	.37	--	.50

(Cont'd.)

*Largest loading of items

Table A-4

Rotated Factor Loadings
Farsi Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.35*	--	--	--	.67
I-43	.50*	.27	--	--	.44
I-44	.51*	.25	--	--	.59
I-45	.37*	.24	.32	--	.70
I-46	.21	.26*	--	--	.34
I-47	.31	.35*	--	.31	.35
I-48	.27	.45*	--	--	.40
I-49	.43*	.33	--	--	.55
I-50	.55*	.32	--	--	.53

*Largest loading of items

Table A-4

Rotated Factor Loadings
Farsi Sample

	I	II	III	IV	Difficulty
II-1	.35*	.32	--	--	.74
II-2	--	.37*	--	--	.63
II-3	.34	--	.37*	--	.83
II-4	.32*	--	.26	--	.68
II-5	.42*	.28	--	--	.69
II-6	.22*	--	--	--	.65
II-7	.25	.28*	--	--	.44
II-8	--	.49*	.27	.36	.12
II-9	--	--	*	--	.44
II-10	.40	.51*	.22	--	.33
II-11	.30	--	.30*	--	.48
II-12	--	--	.24*	--	.26
II-14	--	.29*	.23	.28	.12
II-15	--	.32	.42*	--	.30
II-16	.46*	.24	.24	--	.83
II-17	.43*	.31	.22	.32	.75
II-18	.25*	--	--	--	.67
II-19	--	.23	.24*	--	.54
II-20	.20	--	.22*	--	.71
II-21	.39*	.22	.24	.30	.72
II-22	.35*	.31	.22	.27	.68
II-23	.24	.23	.30*	--	.59
II-24	.25*	.25	.23	--	.44
II-25	--	.33	.34*	--	.57
II-26	--	.35	.42*	--	.36
II-27	.38	.26	.38*	--	.57
II-28	--	--	.31*	--	.61
II-29	.33	.28	.39*	--	.59
II-30	--	.31*	.24	--	.39
II-31	--	--	.20*	--	.39
II-32	--	*	--	--	.21
II-33	.27	.30*	.25	--	.61
II-34	--	--	--	*	.41
II-35	--	.35*	.20	--	.58
II-36	--	.31*	.22	.22	.23
II-37	.26	.39*	.30	--	.38
II-38	--	.22	.25*	--	.33
II-39	--	.23*	--	--	.23
II-40	--	.31*	--	.23	.20

*Largest loading of items

Table A-4

Rotated Factor Loadings
Farsi Sample

	I	II	III	IV	Difficulty
III-1	.26*	--	.25	.21	.82
III-2	.24	--	.36*	--	.74
III-3	.24	--	34*	--	.58
III-4	--	.29	.37*	--	.47
III-5	--	.24	.35*	--	.40
III-6	.26*	--	--	--	.85
III-7	--	--	.37*	--	.74
III-8	.20	--	.38*	--	.56
III-9	.39*	.34	.26	--	.67
III-10	.21	--	.46*	--	.60
III-11	--	--	*	--	.47
III-12	.22*	--	.20	--	.64
III-13	--	.26*	--	--	.38
III-14	--	--	.40*	--	.32
III-15	*	--	--	--	.42
III-16	--	.25	.30*	--	.25
III-17	--	*	--	--	.44
III-18	.24	.40*	--	--	.43
III-19	--	.32*	--	--	.38
III-20	--	.23*	--	--	.35
III-21	--	.43*	--	.23	.29
III-22	--	.33*	.21	--	.37
III-23	--	--	*	--	.32
III-24	--	.20	.33*	--	.21
III-25	--	.33	.34*	--	.23
III-26	--	.32*	.26	--	.26
III-27	--	.20*	.20	--	.24
III-28	--	.37*	--	--	.23
III-29	.22	.39*	--	--	.18
III-30	--	--	--	*	.13
III-31	.43*	.29	.31	--	.57
III-32	.46*	--	.31	--	.79
III-33	--	--	--	*	.33
III-34	.40*	--	.28	--	.78
III-35	.52*	--	.33	--	.83
III-36	.38*	--	.22	--	.61
III-37	.41	--	.44*	--	.64
III-38	.25	--	.42*	--	.47
III-39	--	.21	.30*	.22	.69
III-40	--	--	.26*	--	.56
III-41	.43*	--	.42	--	.62

(Cont'd.)

*Largest loading of items

Table A-4

Rotated Factor Loadings
Farsi Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	.39	--	.40*	--	.68
III-43	.25*	--	.22	.23	.66
III-44	.39*	.27	.37	--	.52
III-45	.21	.24	.29*	--	.45
III-46	.47	--	.53*	--	.62
III-47	.33	.21	.46*	--	.51
III-48	--	--	.23*	--	.47
III-49	--	.23	.27*	--	.50
III-50	.25	--	.28*	--	.36
III-51	.21	--	.43*	--	.36
III-52	.27	.25	.43*	--	.31
III-53	--	--	.32*	--	.48
III-54	.38	--	.39*	--	.57
III-55	.36	--	.44*	--	.64
III-56	--	.30	.30*	--	.55
III-57	--	.31*	.24	--	.37
III-58	.21	.34	.42*	--	.32
III-59	--	.24	.55*	--	.37
III-60	.36	--	.40*	--	.55
Variance	15.52	10.47	9.65	2.78	
Coefficients of Congruence	.990	.899	.908		

*Largest loading of items

Table A-5

Rotated Factor Loadings
Germanic Sample

	I	II	III	IV	Difficulty
I-1	.34*	.27	--	-.21	.97
I-2	.30	.31*	--	-.22	.77
I-3	.40*	--	.32	--	.83
I-4	.37*	--	--	--	.95
I-5	.20	.24*	.23	--	.96
I-6	.23	.30*	--	--	.81
I-7	.34	.44*	--	--	.90
I-8	.42*	.31	--	--	.94
I-9	.32*	.27	--	--	.89
I-10	.37*	--	--	--	.88
I-11	.41*	.27	--	--	.93
I-12	.48*	.37	--	.24	.90
I-13	.52*	--	--	.30	.99
I-14	.48*	.29	--	--	.83
I-15	.50*	--	.25	--	.80
I-16	.39*	.20	--	-.20	.62
I-17	.48*	.31	--	--	.80
I-18	.36*	--	--	--	.60
I-19	.31*	.24	--	--	.83
I-20	.31*	.24	--	--	.72
I-21	.33*	.25	--	--	.94
I-22	.39*	--	--	--	.98
I-23	.47*	.36	--	-.20	.96
I-24	.55*	--	--	--	.86
I-25	.63*	.23	--	--	.95
I-26	.32*	.24	--	--	.96
I-27	.39*	.22	--	-.22	.95
I-28	.46*	.23	--	--	.91
I-29	.23*	--	--	--	.85
I-30	.37*	.27	--	--	.95
I-31	.65*	.25	--	--	.84
I-32	.42*	.28	--	.21	.69
I-33	.52*	.22	--	.33	.83
I-34	.48*	.26	--	--	.89
I-35	.42*	.23	.24	--	.83
I-36	.36	.41*	.32	--	.24
I-37	.21*	.21	--	--	.75
I-38	.24	--	.30*	--	.96
I-39	.47*	.21	.32	--	.59
I-40	--	.35*	--	--	.97
I-41	.48*	--	.23	--	.85

*Largest loading of item

(Cont'd.)

Table A-5

Rotated Factor Loadings
Germanic Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.44*	--	--	--	.95
I-43	.46*	--	--	--	.91
I-44	.22*	--	--	--	.93
I-45	.38*	.33	.31	--	.87
I-46	*	--	--	--	.78
I-47	.51*	.31	.25	--	.86
I-48	.66*	.32	--	--	.83
I-49	.64*	.43	--	--	.93
I-50	.54*	.25	--	--	.95

*Largest loading of item

Table A-5

Rotated Factor Loadings
Germanic Sample

	I	II	III	IV	Difficulty
II-1	.37	.39*	--	--	.92
II-2	--	*	--	--	.84
II-3	.26	.43*	--	.36	.98
II-4	--	.38*	.23	.26	.90
II-5	.20	.35	--	.41*	.98
II-6	.22	.44*	.24	--	.78
II-7	.25	.38*	--	--	.86
II-8	.48*	.36	.23	-.21	.59
II-9	--	.20	.34*	--	.56
II-10	.27	.49*	.29	--	.86
II-11	--	.23*	--	--	.82
II-12	--	.33	.39*	--	.73
II-14	--	--	.55*	-.29	.31
II-15	--	.31	.34*	-.35*	.45
II-16	.22	.48*	.37	--	.98
II-17	.49*	.40	--	.38	.98
II-18	--	.46*	--	--	.92
II-19	--	.35*	.21	.27	.90
II-20	--	.45*	--	--	.91
II-21	.27	.46*	.24	--	.91
II-22	--	.80*	--	--	.91
II-23	--	.42*	.23	.24	.88
II-24	--	.25	.31*	--	.64
II-25	-.23	.36*	.34	--	.81
II-26	--	.46*	.45	--	.81
II-27	.23	.62*	.22	--	.94
II-28	--	.66*	--	--	.92
II-29	.20	.69*	.27	.21	.93
II-30	--	--	.32*	--	.50
II-31	--	.22	.30*	--	.76
II-32	--	.22	.25*	--	.55
II-33	--	.55*	--	--	.92
II-34	.21	.22	.28*	--	.73
II-35	.32	.38*	.23	--	.80
II-36	--	.36*	.31	--	.82
II-37	.20	.49*	.30	--	.79
II-38	--	.28	.45*	--	.62
II-39	--	.36	.45*	--	.55
II-40	--	--	.47*	--	.40

*Largest loading of item

Table A-5

Rotated Factor Loadings
Germanic Sample

	I	II	III	IV	Difficulty
III-1	--	.22	.23*	--	.97
III-2	--	.65*	--	--	.98
III-3	.24	.42*	--	--	.83
III-4	.25	.25	.23	.50*	.94
III-5	--	.32	.27	.33*	.88
III-6	--	.23	--	.24*	.98
III-7	.27*	--	--	.24	.83
III-8	--	.35	.51*	--	.81
III-9	.28	.55*	--	--	.94
III-10	.20	.21	.33*	--	.84
III-11	--	.33*	--	--	.76
III-12	.26	.40*	--	--	.74
III-13	--	.26	.35*	--	.79
III-14	.24	--	.51*	--	.90
III-15	--	--	.23*	--	.37
III-16	.29	--	.39*	-.22	.56
III-17	--	--	.23*	--	.59
III-18	.23	.31	.24	-.42*	.71
III-19	.28	.42*	.29	--	.76
III-20	--	.33*	--	--	.56
III-21	.34	.49*	.26	--	.76
III-22	--	.45*	.26	--	.76
III-23	--	--	.26*	--	.45
III-24	--	.25	.39*	--	.37
III-25	--	.33*	.43*	--	.67
III-26	--	--	.62*	--	.62
III-27	--	--	.30*	--	.43
III-28	--	.27	.29*	--	.66
III-29	--	.34	.44*	--	.56
III-30	--	.27*	--	-.20	.12
III-31	.45*	.42	.21	.33	.97
III-32	--	.39*	--	--	.98
III-33	.25	--	.20	.26*	.83
III-34	.36*	.21	--	--	.97
III-35	--	--	.51*	--	.98
III-36	.37*	.33	.28	.36	.94
III-37	--	.38*	.31	--	.97
III-38	.23	.32	.42*	--	.77
III-39	--	.27	.28*	.22	.94
III-40	--	--	--	*	.90
III-41	.32	.36	.41*	.40	.94

*Largest loading of item

(Cont'd.)

Table A-5

Rotated Factor Loadings
Germanic Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	--	.27	.24	.38*	.94
III-43	.23*	--	--	--	.88
III-44	.22	.34	.50*	.40	.90
III-45	--	.26	.28*	--	.75
III-46	.48	.39	.50*	.31	.97
III-47	.32	.36*	.36	--	.88
III-48	--	.31	.37*	--	.66
III-49	--	--	.43*	.20	.93
III-50	--	.25	.40*	--	.88
III-51	.32	.21	.36*	.26	.81
III-52	.33	.20	.47*	--	.70
III-53	.20	.21	.36*	.27	.85
III-54	.38	.30	.44	.49*	.92
III-55	.46	.28	.55*	.20	.93
III-56	.36	.29	.48*	.37	.93
III-57	--	--	.54*	--	.78
III-58	--	.38	.55*	.29	.85
III-59	--	--	.71*	.37	.89
III-60	.36	.47*	.42	.23	.90
Variance	14.08	15.62	12.44	5.34	
Coefficients of Congruence	.965	.957	.924		

*Largest loading of item

Table A-6

Rotated Factor Loadings
Japanese Sample

	I	II	III	IV	Difficulty
I-1	.51*	--	--	--	.91
I-2	.22*	--	--	--	.73
I-3	.40*	--	.20	--	.69
I-4	.35*	--	--	--	.56
I-5	.28*	--	--	--	.83
I-6	.42*	.25	--	--	.61
I-7	.37*	.23	.21	--	.61
I-8	.51*	--	--	--	.86
I-9	.49*	.23	.25	--	.72
I-10	--	.20*	--	--	.37
I-11	.45*	.35	--	--	.60
I-12	.46*	--	--	--	.54
I-13	.42*	.32	--	--	.74
I-14	.43*	.29	--	--	.52
I-15	.51*	.29	--	--	.50
I-16	*	--	--	--	.37
I-17	.48*	.23	--	--	.51
I-18	.23*	--	--	--	.33
I-19	--	*	--	--	.56
I-20	.21	.22*	--	--	.35
I-21	--	*	--	--	.83
I-22	.52*	--	--	-.27	.91
I-23	.36*	.30	--	--	.86
I-24	.38*	.22	--	--	.70
I-25	.63*	--	--	--	.77
I-26	.57*	--	--	--	.88
I-27	.59*	.23	--	--	.70
I-28	.46*	--	--	--	.67
I-29	.54*	.21	--	--	.58
I-30	.38*	.29	--	--	.41
I-31	.46*	.21	--	--	.42
I-32	.50*	.31	--	--	.31
I-33	.44*	--	--	--	.50
I-34	.47*	--	--	--	.51
I-35	.32*	--	--	--	.44
I-36	.47*	.22	.21	--	.61
I-37	--	*	--	--	.65
I-38	.37	.28	.48*	-.28	.85
I-39	*	--	--	--	.22
I-40	.41	.27	.41*	-.31	.83
I-41	.32*	--	.30	--	.65

(Cont'd.)

*Largest loading of items

Table A-6

Rotated Factor Loadings
Japanese Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.31*	--	--	--	.84
I-43	.57*	.25	--	--	.61
I-44	.54*	.23	--	--	.60
I-45	.47*	.32	.39	-.26	.81
I-46	.37*	.24	--	--	.31
I-47	.39*	.35	--	--	.50
I-48	.41*	.33	--	--	.33
I-49	.49*	.34	--	--	.56
I-50	.42*	.28	--	--	.54

*Largest loading of items

Table A-6

Rotated Factor Loadings
Japanese Sample

	I	II	III	IV	Difficulty
II-1	.32	.33*	.25	--	.79
II-2	--	.36*	.20	--	.75
II-3	.25	.39*	.33	-.24	.87
II-4	.27	.43*	.20	--	.77
II-5	.31	.38*	.21	--	.80
II-6	--	--	--	*	.57
II-7	--	.31*	--	--	.63
II-8	.25	.26	--	.40*	.17
II-9	--	--	--	*	.42
II-10	.34	.43*	--	--	.36
II-11	.22	.41*	--	--	.46
II-12	--	.27*	.26	--	.28
II-14	--	.46*	--	--	.18
II-15	--	.41	.46*	--	.45
II-16	.27	.39*	.37	--	.86
II-17	.34*	.30	--	-.26	.87
II-18	--	.34*	--	--	.74
II-19	--	.38*	--	-.25	.74
II-20	--	.30*	.21	--	.83
II-21	--	.53*	.28	-.27	.78
II-22	.27	.47*	.26	--	.80
II-23	--	.23*	--	--	.62
II-24	.25	.30*	.27	--	.58
II-25	.24	.45*	.31	-.35	.79
II-26	--	.35*	.25	--	.62
II-27	.22	.44*	.35	--	.66
II-28	--	--	.28*	--	.75
II-29	.22	.42*	.23	--	.70
II-30	--	.43*	.20	--	.53
II-31	--	.33*	.22	--	.50
II-32	--	.23*	--	--	.35
II-33	--	.43*	--	--	.66
II-34	--	.32*	--	--	.48
II-35	--	.40*	--	--	.78
II-36	--	--	--	.30*	.19
II-37	.25	.46*	--	--	.38
II-38	--	.40*	--	--	.44
II-39	--	.27*	--	--	.19
II-40	--	.27	.31*	--	.28

*Largest loading of items

Table A-6

Rotated Factor Loadings
Japanese Sample

	I	II	III	IV	Difficulty
III-1	.27	--	.37*	--	.87
III-2	.25	--	.45*	--	.85
III-3	--	.27*	.27	--	.65
III-4	.23	--	.37*	--	.72
III-5	--	.33*	.38*	--	.63
III-6	.25	--	25*	--	.87
III-7	--	--	.24*	--	.88
III-8	--	--	.36*	--	.73
III-9	.30	.33	.39*	--	.62
III-10	--	--	.47*	--	.60
III-11	--	.26	.34*	--	.72
III-12	.31*	.21	.23	--	.64
III-13	--	--	.44*	--	.58
III-14	--	--	.25*	--	.40
III-15	--	--	--	.20*	.33
III-16	--	.20	.53*	--	.35
III-17	--	.21	.31*	--	.57
III-18	.29*	.28	--	--	.40
III-19	--	.27	.34*	--	.62
III-20	--	*	--	--	.45
III-21	.24	.32*	.23	--	.39
III-22	--	--	.37*	--	.87
III-23	--	--	.22*	--	.25
III-24	--	.21	.37*	--	.41
III-25	--	.31	.60*	--	.42
III-26	--	.31	.43*	--	.27
III-27	--	--	*	--	.61
III-28	--	.25	.48*	--	.28
III-29	--	--	*	--	.26
III-30	--	--	--	.27*	*
III-31	.36	.42*	.38	--	.70
III-32	.38*	--	.30	--	.84
III-33	--	.31*	--	.20	.41
III-34	.25	.28*	--	--	.85
III-35	.39*	--	.25	--	.95
III-36	.27	.28*	--	--	.69
III-37	.25	.39	.46*	--	.69
III-38	.25	.32	.43*	--	.62
III-39	--	.31	.38*	--	.88
III-40	--	.21*	--	--	.71
III-41	.33	.46*	.22	--	.72

(Cont'd.)

*Largest loading of items

Table A-6

Rotated Factor Loadings
Japanese Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	.25	.38*	.23	--	.76
III-43	--	.52*	--	--	.83
III-44	--	.59*	.35	--	.60
III-45	.30*	.29	.20	--	.52
III-46	.35	.45	.54*	-.28	.78
III-47	--	.45*	.32	--	.59
III-48	--	.35*	.26	--	.58
III-49	.26	.32	.34*	--	.65
III-50	.26	.41*	.22	--	.62
III-51	.28	.45*	.37	--	.46
III-52	.26	.41*	.31	.21	.36
III-53	--	.30*	--	--	.50
III-54	.30	.46*	.36	--	.75
III-55	.23	.56*	.33	--	.70
III-56	.20	.53*	.36	--	.67
III-57	--	.43*	--	--	.48
III-58	.26	.53*	.40	--	.46
III-59	--	.54*	.36	--	.59
III-60	.27	.49*	.38	--	.70
Variance	13.12	14.24	9.97	2.59	
Coefficient of Congruence	.983	.938	.954		

*Largest loading of items

Table A-7

Rotated Factor Loadings
Spanish Sample

	I	II	III	IV	Difficulty
I-1	.62*	.26	.21	--	.90
I-2	.32*	.31	--	--	.66
I-3	.45*	.25	--	--	.79
I-4	.35*	.30	.33	.29	.86
I-5	.52*	.30	.30	.27	.85
I-6	.56*	.25	--	--	.68
I-7	.30	.36*	.21	--	.75
I-8	.45*	.37	--	--	.69
I-9	.44*	--	.24	--	.67
I-10	.50*	.38	--	--	.61
I-11	.61*	.42	--	--	.73
I-12	.46	.51*	--	--	.67
I-13	.41*	.33	.23	--	.84
I-14	.44	.45*	--	--	.71
I-15	.54*	.44	--	--	.66
I-16	.30	.47*	--	--	.36
I-17	.52*	.43	--	--	.61
I-18	--	.47*	--	--	.39
I-19	.22	.32*	--	--	.54
I-20	.26	.30*	--	--	.46
I-21	.33*	.31	--	--	.83
I-22	.63*	--	--	--	.92
I-23	.53*	.23	.28	.23	.89
I-24	.44*	--	.35	--	.80
I-25	.56*	.24	--	--	.86
I-26	.54*	.31	.24	--	.85
I-27	.68*	.21	--	--	.85
I-28	.41*	--	.22	--	.77
I-29	.42*	.24	--	--	.73
I-30	.64*	.47	--	--	.66
I-31	.51*	.38	--	--	.67
I-32	.58*	.42	--	--	.54
I-33	.43	.51*	--	--	.62
I-34	.39*	.36	--	--	.71
I-35	.36	.45*	--	--	.49
I-36	.60*	.26	.34	--	.80
I-37	--	*	--	--	.63
I-38	.46*	--	.38	--	.93
I-39	.38*	.35	.21	--	.36
I-40	.51*	--	.36	--	.90
I-41	.38*	--	.36	--	.77

(Cont'd.)

*Largest loading of items

Table A-7

Rotated Factor Loadings
Spanish Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.38*	--	.22	--	.80
I-43	.54*	.26	.20	--	.70
I-44	.50*	.30	--	--	.82
I-45	.53*	.29	.26	--	.79
I-46	.27	.39*	--	--	.46
I-47	.33	.47*	.24	--	.60
I-48	.49*	.45	--	--	.56
I-49	.60*	.42	--	--	.74
I-50	.52*	.33	--	--	.80

*Largest loading of items

Table A-7

Rotated Factor Loadings
Spanish Sample

	I	II	III	IV	Difficulty
II-1	.44*	.35	--	--	.82
II-2	--	--	.23*	--	.71
II-3	.42*	.29	.30	--	.88
II-4	--	.27	.29*	--	.81
II-5	.34	.44*	.31	.24	.86
II-6	.27	.34*	.30	--	.80
II-7	.39*	.34	--	--	.73
II-8	.27	.56*	--	--	.31
II-9	--	--	*	--	.41
II-10	.44	.55*	.20	--	.62
II-11	.37	.22	.37*	--	.72
II-12	--	.60*	.29	--	.44
II-14	--	.52*	.28	--	.17
II-15	--	.48*	.35	--	.51
II-16	.40	.38*	.27	.40	.92
II-17	.44	.48*	--	.24	.80
II-18	.20	.33*	.30	--	.83
II-19	.20	.34*	.26	--	.61
II-20	--	.29	.35*	.24	.85
II-21	.33	.35*	.29	--	.73
II-22	.47*	.42	.29	--	.70
II-23	.25	.44*	.20	--	.70
II-24	.26	.43*	.28	--	.71
II-25	--	.34*	.32	--	.58
II-26	--	.49*	.42	--	.61
II-27	.37	.44*	.27	--	.70
II-28	--	.49*	.29	--	.80
II-29	.28	.57*	--	--	.72
II-30	--	.45*	--	--	.40
II-31	--	.24	.26*	--	.57
II-32	--	--	.22*	--	.55
II-33	.24	.52*	.30	--	.74
II-34	--	.35*	--	--	.49
II-35	--	.44*	--	--	.55
II-36	.35	.51*	.21	--	.60
II-37	.27	.60*	.28	--	.47
II-38	--	.29*	--	--	.35
II-39	--	.35*	--	.24	.46
II-40	--	.48*	.26	--	.27

(Cont'd.)

*Largest loading of items

Table A-7

Rotated Factor Loadings
Spanish Sample

	I	II	III	IV	Difficulty
III-1	--	--	.21*	--	.96
III-2	.23	.42*	--	--	.81
III-3	.24	.40*	--	--	.71
III-4	.44*	--	.37	--	.86
III-5	.20	--	.59*	--	.85
III-6	--	.20	.42*	--	.94
III-7	.31	.39*	.24	--	.77
III-8	--	--	.50*	--	.96
III-9	.48	.50*	.20	--	.72
III-10	--	--	.32*	--	.73
III-11	--	.33*	.26	--	.70
III-12	.35	.44*	.24	--	.59
III-13	--	--	.44*	--	.94
III-14	.24	.33	.51*	.33	.98
III-15	--	.20*	--	--	.44
III-16	.22	.23	.38*	--	.60
III-17	--	--	--	*	.68
III-18	.39	.62*	--	--	.56
III-19	--	--	.32*	--	.84
III-20	--	.24	.30*	--	.51
III-21	.29	.59*	--	--	.34
III-22	--	--	.52*	--	.96
III-23	--	--	.41*	--	.70
III-24	--	-.20	.57*	--	.82
III-25	.34	.24	.59*	--	.90
III-26	--	.21	.25*	-.23	.58
III-27	--	--	.23*	--	.34
III-28	.20	.25	.27*	--	.62
III-29	.27	.21	.40*	-.32	.50
III-30	--	--	*	--	.20
III-31	.29	.31	.56*	--	.91
III-32	.25	--	.52*	--	.96
III-33	.28*	.21	.22	--	.57
III-34	.28	--	.42*	--	.90
III-35	--	--	.47*	--	.94
III-36	.25	--	.30*	--	.74
III-37	.31	.20	.49*	--	.88
III-38	.24	.33	.38*	--	.73
III-39	--	.24	.51*	.20	.84
III-40	--	--	.30*	--	.85
III-41	.35	.29	.53*	--	.81

(Cont'd.)

*Largest loading of items

Table A-7

Rotated Factor Loadings
Spanish Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	.20	--	.53*	--	.92
III-43	--	--	.42*	.21	.85
III-44	.25	.31	.46*	--	.76
III-45	.22	--	.41*	--	.66
III-46	.31	.34	.54*	--	.84
III-47	.28	.25	.48*	--	.80
III-48	.23	.26	.33*	--	.70
III-49	.31	.29	.51*	--	.81
III-50	.37	.20	.39*	--	.77
III-51	.24	--	.57*	--	.69
III-52	.33	.36	.53*	--	.56
III-53	.26	.20	.44*	--	.75
III-54	.44	--	.57*	--	.86
III-55	.43	.28	.55*	--	.85
III-56	.45	.23	.46*	--	.80
III-57	.29	.49*	.26	--	.62
III-58	.33	.50*	.37	--	.50
III-59	.36	.28	.46*	-.27	.79
III-60	.37	.38	.42*	-.21	.72
Variance	18.34	17.50	14.63	2.45	
Coefficient of Congruence	.981	.923	.900		

*Largest loading of items

Table A-8

Rotated Factor Loadings
Total Sample

	I	II	III	IV	Difficulty
I-1	.52*	.23	--	--	.90
I-2	.39*	--	--	--	.65
I-3	.40*	.26	--	--	.71
I-4	.49*	--	--	--	.74
I-5	.32*	.29	--	.29	.85
I-6	.47*	--	--	--	.62
I-7	.40*	.37	--	--	.76
I-8	.54*	.22	--	--	.74
I-9	.25	.27*	--	.22	.74
I-10	.44	.24	--	--	.55
I-11	.59*	.28	--	--	.63
I-12	.47*	.38	--	--	.61
I-13	.41*	.41	--	--	.82
I-14	.39*	.33	.24	--	.63
I-15	.54*	--	--	--	.55
I-16	.36*	--	--	--	.30
I-17	.51*	.31	--	--	.58
I-18	.24*	--	--	--	.40
I-19	.31*	--	.23	--	.52
I-20	.21	.20	.25*	--	.39
I-21	.47*	--	.27	--	.75
I-22	.49*	.32	--	.24	.93
I-23	.54*	--	--	--	.85
I-24	.46*	--	--	--	.65
I-25	.52*	--	--	--	.79
I-26	.53*	.33	.22	.21	.81
I-27	.64*	--	--	--	.76
I-28	.46*	--	--	--	.69
I-29	.53*	.24	--	.23	.66
I-30	.54*	.30	--	--	.63
I-31	.61*	--	.22	--	.52
I-32	.59*	.27	--	--	.46
I-33	.49*	.31	.24	--	.60
I-34	.50*	.30	--	--	.63
I-35	.39*	.29	--	--	.53
I-36	.49*	--	.29	--	.68
I-37	--	--	*	--	.56
I-38	.46*	.27	.34	.29	.84
I-39	.37	--	.38*	--	.28
I-40	.39*	.22	.32	.24	.83
I-41	.29	.31*	.27	--	.65

(Cont'd.)

*Largest loading of item

Table A-8

Rotated Factor Loadings
Total Sample

(Cont'd.)

	I	II	III	IV	Difficulty
I-42	.42*	--	.28	--	.75
I-43	.64*	--	.28	--	.57
I-44	.57*	--	--	--	.72
I-45	.35	.39*	--	--	.81
I-46	.40*	--	--	--	.38
I-47	.47*	--	.37	--	.45
I-48	.58*	.23	.25	--	.45
I-49	.56*	.40	--	--	.70
I-50	.43*	.30	.25	--	.71

*Largest loading of item

Table A-8

Rotated Factor Loadings
Total Sample

	I	II	III	IV	Difficulty
II-1	.37	.38*	.27	--	.80
II-2	--	.27	.35*	--	.79
II-3	.33	.50*	.28	.23	.89
II-4	--	.39*	.21	--	.82
II-5	.41*	.36	.25	.20	.82
II-6	.22	.35	.23	--	.68
II-7	.20	.34*	.25	--	.69
II-8	--	.61*	--	.29	.47
II-9	--	.24	.29*	--	.55
II-10	.36	.41	.47*	--	.54
II-11	.25	.31	.44*	--	.60
II-12	--	.43*	.42	.20	.45
II-14	--	.50*	.36	--	.36
II-15	--	.52*	.39	--	.52
II-16	.35	.69*	--	.22	.91
II-17	.35	.50*	--	.23	.87
II-18	.27	.44*	--	--	.75
II-19	.22	.31	.40*	--	.65
II-20	--	--	.33*	--	.77
II-21	.22	.39*	.26	.30	.80
II-22	.30	.64*	.21	--	.82
II-23	.22	.33*	.30	--	.68
II-24	.23	.37*	.23	.20	.60
II-25	.22	.30	.34*	.35	.70
II-26	.26	.29	.52*	.27	.60
II-27	.25	.59	.27	--	.71
II-28	--	.43*	.31	--	.77
II-29	.27	.52*	.35	--	.77
II-30	--	.51*	.31	--	.56
II-31	--	.33	.30	--	.54
II-32	--	--	.30*	--	.40
II-33	.27	.55*	.21	--	.76
II-34	--	.32*	.25	--	.56
II-35	.25	.38*	.27	--	.69
II-36	.21	.42*	.22	--	.46
II-37	.25	.30	.39*	--	.43
II-38	--	.26	.34*	--	.42
II-39	--	.27	.31*	--	.37
II-40	--	.41	.42*	--	.38

*Largest loading of item

Table A-8

Rotated Factor Loadings
Total Sample

	I	II	III	IV	Difficulty
III-1	.28	--	.39*	--	.80
III-2	.20	.55*	.21	--	.87
III-3	--	.58*	--	--	.76
III-4	.28	.36	.37*	--	.70
III-5	--	.37	.48*	--	.66
III-6	.37*	--	.32	--	.79
III-7	.24*	--	--	--	.80
III-8	--	.23	.46*	--	.71
III-9	.42	.42*	.28	--	.72
III-10	--	.32	.35*	--	.68
III-11	--	.26	.33*	--	.67
III-12	.29	.41*	--	--	.71
III-13	--	.42*	.42	.22	.69
III-14	--	.34	.54*	--	.62
III-15	--	--	.36*	--	.42
III-16	--	.53*	.36	--	.51
III-17	--	--	.21*	--	.54
III-18	.26	.44*	.40	--	.58
III-19	.26	--	.52*	--	.56
III-20	--	.35	.36	--	.54
III-21	.22	.39*	--	--	.47
III-22	--	--	.52	--	.67
III-23	--	.37	.50*	.20	.49
III-24	--	.21	.62*	--	.47
III-25	.20	--	.60*	.20	.49
III-26	--	.42	.48*	--	.53
III-27	--	.33	.38*	--	.48
III-28	--	--	.63*	--	.37
III-29	--	.31	.53*	--	.39
III-30	--	.22	.34*	.23	*
III-31	.36	.35	.52*	--	.75
III-32	.31	.49*	.20	--	.90
III-33	.31*	.27	.21	--	.54
III-34	.30*	.26	.28	--	.83
III-35	.42*	.29	.24	.24	.91
III-36	.25	.20	.33*	.24	.73
III-37	.24	.44*	.40	--	.79
III-38	.25	.37*	.27	.24	.67
III-39	.29	--	.41*	.40	.74
III-40	--	--	.29	--	.74
III-41	.39*	.37	.38	.22	.74

(Cont'd.)

*Largest loading of item

Table A-8

Rotated Factor Loadings
Total Sample

(Cont'd.)

	I	II	III	IV	Difficulty
III-42	.29	.28	.36*	--	.79
III-43	.20	.35*	--	.34	.81
III-44	.31	.33	.48*	--	.68
III-45	.30	.22	.34*	--	.56
III-46	.34	.61*	.30	.29	.8[illegible]
III-47	.31	.40	.49*	--	.69
III-48	.24	--	.32*	.26	.51
III-49	.22	.46*	.32	--	.76
III-50	--	.27	.29*	--	.69
III-51	.29	.33	.43*	.23	.58
III-52	.29	.40	.43*	.23	.50
III-53	--	.27	.34*	--	.67
III-54	.28	.36	.37*	.21	.78
III-55	.33	.38	.45*	.33	.75
III-56	.28	.35*	.47	--	.75
III-57	.24	.31	.37*	--	.53
III-58	.31	.41	.52*	--	.55
III-59	.38	.31	.49*	.22	.62
III-60	.26	.41*	.33	.22	.69
Variance	16.96	17.09	15.28	3.63	
Coefficients of Congruence	.986	.972	.978		

*Largest loading of item

Appendix B

Factors Extension Correlation Matrices

Table B-1

Factor Extension Matrix for African Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)			-.17	
Birth year			-.31	
Reason for Taking TOEFL:				
Undergraduate Study	-.13		-.23	-.17
Graduate Study	.15		.27	.18
Other Study				
Professional Licensing				
Employment Requirement				
Other				
Number of Previous TOEFLS				
Studying for Degree				
Listening Comprehension (Number Correct)	.75	.18	.24	
Structure & Written Expression (Number Correct)		.57	.31	.13
Reading Comp. & Vocab. (Number Correct)	.31	.39	.50	.30

Note. Loadings less than .10 have been omitted.

Table B-2

Factor Extension Matrix for Arabic Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)				
Birth year	.21	.13	-.17	-.27
Reason for Taking TOEFL:				
Undergraduate Study	.14		-.15	-.32
Graduate Study	-.19			.22
Other Study				-.11
Professional Licensing			.12	.20
Employment Requirement	.10			
Other				
Number of Previous TOEFLS				
Studying for Degree	-.15			
Listening Comprehension (Number Correct)	.62	.36	.26	
Structure & Written Expression (Number Correct)	.20	.61	.35	
Reading Comp. & Vocab. (Number Correct)	.21	.43	.62	

Note: Loadings less than .10 have been omitted.

Table B-3

Factor Extension Matrix for Chinese Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)				
Birth year			-.18	
Reason for Taking TOEFL:				
Undergraduate Study		.11		
Graduate Study				
Other Study				
Professional Licensing				
Employment Requirement				
Other				
Number of Previous TOEFLS				
Studying for Degree				
Listening Comprehension (Number Correct)	.66	.35	.18	-.16
Structure & Written Expression (Number Correct)	.25	.51	.43	
Reading Comp. & Vocab. (Number Correct)	.20	.56	.49	

Note. Loadings less than .10 have been omitted.

Table B-4

Factor Extension Matrix for Farsi Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)				
Birth year		.29		
Reason for Taking TOEFL:				
Undergraduate Study		.27	-.21	.10
Graduate Study		-.27	.23	-.16
Other Study				
Professional Licensing				
Employment Requirement				
Other				
Number of Previous TOEFLS				
Studying for Degree				
Listening Comprehension (Number Correct)	.64	.45	.10	
Structure & Written Expression (Number Correct)	.33	.49	.43	.17
Reading Comp. & Vocab. (Number Correct)	.32	.33	.63	

Note. Loadings less than .10 have been omitted.

Table B-5

Factor Extension Matrix for Germanic Sample

Varible	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)		.12	-.10	
Birth year		.10	-.15	
Reason for Taking TOEFL:				
Undergraduate Study		.11	-.14	
Graduate Study			.18	
Other Study				
Professional Licensing				
Employment Requirement				
Other				
Number of Previous TOEFLS				
Studying for Degree				.11
Listening Comprehension (Number Correct)	.58	.26	.18	.17
Structure & Written Expression (Number Correct)		.48	.46	.28
Reading Comp. & Vocab. (Number Correct)	.15	.33	.60	

Note. Loadings less than .10 have been omitted.

Table B-6

Factor Extension Matrix for Japanese Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)	.14		-.27	
Birth year			-.25	
Reason for Taking TOEFL:				
Undergraduate Study		-.10	-.21	
Graduate Study		.14	.21	
Other Study	-.23	-.12		
Professional Licensing				
Employment Requirement				
Other				
Number of Previous TOEFLS			.10	.10
Studying for Degree			.11	
Listening Comprehension (Number Correct)	.68	.35	.14	
Structure & Written Expression (Number Correct)	.21	.67	.28	
Reading Comp. & Vocab. (Number Correct)	.23	.51	.53	

Note. Loadings less than .10 have been omitted.

Table B-7

Factor Extension Matrix for Spanish Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)		.16	-.14	.10
Birth year	.15	.22	-.21	.16
Reason for Taking TOEFL:				
Undergraduate Study	.11	.26	-.26	.12
Graduate Study		-.17	.20	
Other Study		-.10		
Professional Licensing				
Employment Requirement				
Other				
Number of Previous TOEFLS				
Studying for Degree		.10	-.11	
Listening Comprehension (Number Correct)	.62	.46		
Structure & Written Expression (Number Correct)	.26	.68	.23	-.11
Reading Comp. & Vocab. (Number Correct)	.32	.40	.49	-.19

Note. Loadings less than .10 have been omitted.

Table B-8

Factor Extension Matrix for Total Sample

Variable	Factor I	II	III	IV
Sex (1 = Male, 0 = Female)			-.10	
Birth year	.15		-.24	
Reason for Taking TOEFL:				
Undergraduate Study			-.21	-.10
Graduate Study		-.10	.26	.17
Other Study			-.13	
Professional Licensing		.11		-.11
Employment Requirement				
Other				
Number of Previous TOEFLS				.22
Studying for Degree				
Listening Comprehension (Number Correct)	.69	.25	.28	
Structure & Written Expression (Number Correct)	.23	.54	.46	
Reading Comp. & Vocab. (Number Correct)	.26	.40	.62	

Note. Loadings less than .10 have been omitted.

TOEFL
Research Reports

The Performance of Native Speakers of English on the Test of English as a Foreign Language: Clark, John L.D. Report 1. November 1977.

Discusses the results of the administration of TOEFL to native speakers of English just prior to their graduation from a college-preparatory high school program. Total test score distributions were highly negatively skewed, reinforcing findings of earlier studies that TOEFL is not psychometrically appropriate for discriminating among native speakers of English with respect to English language competence.

An Evaluation of Alternative Item Formats for Testing English as a Foreign Language: Pike, Lewis W. Report 2. June 1979.

Describes an extensive research study conducted from 1972 to 1974 that was designed to explore possible changes in the format and content of TOEFL. Questions of validation, criterion selection, and content specifications were investigated. The report includes the results of these findings and discusses the implications for TOEFL content specifications and internal structure. This study contributed to the restructuring of TOEFL beginning in 1976.

The Performance of Non-Native Speakers of English on TOEFL and Verbal Aptitude Tests: Angelis, Paul J.; Swinton, Spencer S.; and Cowell, William R. Report 3. October 1979.

Gives the results of a study in which 400 graduate and undergraduate applicants took TOEFL, the GRE Verbal or the SAT Verbal, and the Test of Standard Written English (TSWE). Included in the report are comparative data on performance across tests and interpretive information on how combined test results might best be used in the admission process.

An Exploration of Speaking Proficiency Measures in the TOEFL Context: Clark, John L.D., and Swinton, Spencer S. Report 4. October 1979.

Describes a three-year study involving the development and experimental administration of test formats and item types aimed at measuring the English-speaking proficiency of nonnative speakers. Factor analysis and other techniques were used to identify subsets of item formats and individual items having satisfactory correlations with the Foreign Service Institute criterion interview administered to the test subjects. The results were grouped into a prototype "Test of Spoken English."

The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language: Powers, Donald E. Report 5. December 1980.

Summarizes analyses indicating performance of 6,000 nonnative speakers of English on TOEFL and GMAT. In addition to comparisons between native and nonnative speakers, data are included showing performance by language background. A variety of analyses support the basic differences in the two tests by showing expected GMAT verbal scores for various levels of TOEFL scores.

Factor Analysis of the Test of English as a Foreign Language for Several Language Groups: Powers, Donald E., and Swinton, Spencer S. Report 6. December 1980.

Provides evidence from a set of exploratory analytical techniques that three major factors underlie performance on TOEFL. Some support is also found for concluding that these factors may be interpreted differently for several language groups. The report discusses implications for making inferences based on TOEFL subscores and considerations for future test development.

The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings: Clark, John L.D., and Swinton, Spencer S. Report 7. December 1980.

Presents the results of a study that examined the performance of foreign teaching assistants on the Test of Spoken English in relation to their classroom performance as judged by students. Also includes, for purposes of comparison, data showing performance of the same groups of teaching assistants on the Foreign Service oral interview and on TOEFL. Based on the analyses conducted in the study, TSE is shown to be a valid predictor of language abilities for nonnative English-speaking graduate teaching assistants.

Effects of Item Disclosure on TOEFL Performance: Angelis, Paul J.; Hale, Gordon A.; and Thibodeau, Lawrence A. Report 8. December 1980.

Reports the findings of a study designed to examine the effects of performance on TOEFL when a subset of items have been disclosed prior to an administration. Based on data from 16 intensive English training programs, the results indicate significant increases in performance in proportion to the number of items made available to students. Details are provided showing separate results by language group and by item type.

The above reports are currently available. Other research reports are planned. For further information about any of the TOEFL Research Reports, write to:

TOEFL Program Office
Box 899
Princeton, NJ 08541, USA